



CLASSIFICATION ET FORÊTS ALÉATOIRES: APPLICATION À  
L'AIDE À LA DÉCISION CHIRURGICALE DU GENOU PAR  
ARTHROPLASTIE

Mémoire présenté comme exigence partielle  
de la maîtrise en informatique

Par Mohamed Kazim MOUDACHIROU

Juillet 2017

<http://r-libre.teluq.ca/1431>



## Résumé

Ce mémoire s'inscrit dans le cadre de l'apprentissage statistique des données pour le développement de système d'aide à la décision. Plus spécifiquement, il est consacré à l'étude des forêts aléatoires qui constituent l'un des plus récents algorithmes de la famille des méthodes de classification des formes. L'objectif étant d'utiliser les forêts aléatoires, qui utilisent le vote majoritaire uniforme d'arbres de décision, pour classer des patients ayant une arthrose au genou selon deux classes : une classe de sujets chirurgicaux pour une arthroplastie du genou et une classe de sujets non chirurgicaux. La classification se base sur les données cinématiques du genou du patient enregistrées dans les plans suivants : frontal, sagittal et transversal. Ces données correspondent respectivement au mouvement tibiofémoral d'abduction/adduction, de flexion/extension et de rotation interne et externe durant une tâche de marche normale. Quatre-vingt-neuf (89) sujets dont quarante-huit (48) chirurgicaux et quarante-et-un (41) patients non chirurgicaux ont participé à ce projet de recherche. Les données initialement fusionnées, sont scindées en deux jeux de données dont le premier a servi à l'échantillon d'entraînement en vue de concevoir un modèle décisionnel d'aide à la décision de l'arthroplastie du genou. La procédure de classement suit un cheminement simple qui est formé par un test concernant les prévisions. Le patient à classer est dirigé selon la valeur de son attribut vers la branche adéquate, pour y être soumis le cas échéant à un nouveau test, et ce jusqu'à atteindre une feuille terminale. L'attribution de l'étiquette de classe est effectuée à travers un vote majoritaire. L'objectif est d'utiliser

ce modèle pour décider à quelle classe affecter de futures (nouvelles) données. Le second jeu de données ou échantillon de test a servi à identifier les sujets chirurgicaux ou non-chirurgicaux avec l'étiquette de classe attribuée aux participants via un vote à majorité simple basé sur le modèle réalisé avec le sous-ensemble de l'échantillon d'apprentissage. Cette identification nous a permis par la suite, de valider notre modèle décisionnel par une matrice de confusion et des tests statistiques d'hypothèses des forêts aléatoires en termes de sensibilité, spécificité et le taux de classification correcte.

**Mots clés :** Classification automatique, Forage des données, Forêts aléatoires, Arbre de décision, Vote majoritaire, Regroupement, Aide au diagnostic, Gonarthrose, Asymptomatique, Genou, Biomécanique du genou, cinématique du genou, cycle de marche, Classification, partitionnement, classification automatique, Système décisionnel, Arthroplastie du genou.



## **Abstract**

This work is part of the statistical learning and is devoted to the study of random forests which is one of the most recent algorithms of the family of methods of classification of the forms which uses the uniform majority vote of trees Decision making of surgical subjects and non-surgical subjects based on kinematic data recorded in the frontal, sagittal and transverse planes. These data correspond respectively to the abduction/adduction, flexion/extension and internal and external rotation during the normal walking cycle. Eighty-nine (89) subjects, including forty-eight (48) patients and forty-one (41) healthy patients, participated in this study. Initially merged data are split into two datasets, the first of dataset was used in the training sample to develop a decision model to aid diagnosis of knee arthroplasty. The classification procedure follows a simple path that is formed by a test concerning the forecasts. The patient to be classified is directed according to the value of his attribute towards the appropriate branch, to be subjected to a new test if necessary, until it reaches a terminal sheet. The class label is awarded by a majority vote. The goal is to use this model to decide which class will be assigned to a future (new) data. The second dataset or test sample was used to identify diseased or healthy subjects with the class label assigned to the participants via a simple majority vote based on the model performed with the subset of the learning sample. This identification enabled us subsequently to validate our decisional model by a confusion matrix and statistical tests of random forest hypotheses in terms of sensitivity, specificity and the correct classification rate.

**Keywords:** Forms classification, Data mining, Random forest, Decision tree, Majoritarian voting, Grouping, Diagnostic aid, Knee osteoarthritis, Asymptomatic, Knee, Knee biomechanics, Knee kinematics, walking cycle, Classification, partitioning, automatic classification, decisional system, knee arthroplasty.

## TABLE DES MATIÈRES

<b>TABLE DES MATIÈRES .....</b>	<b>XV</b>
<b>LISTE DES TABLEAUX .....</b>	<b>XVII</b>
<b>LISTE DES FIGURES .....</b>	<b>XVIII</b>
<b>DE TOUT CŒUR, MERCI A TOUS!.....</b>	<b>XVIII</b>
1.1. <i>Contexte, justification et problématique .....</i>	<i>4</i>
1.2. <i>Objectifs .....</i>	<i>5</i>
<b>CHAPITRE 2.....</b>	<b>7</b>
<b>LE GENOU ET SA BIOMECHANIQUE .....</b>	<b>7</b>
1.    L'ANATOMIE DU GENOU .....	7
1.1. <i>Le fémur .....</i>	<i>8</i>
1.2. <i>Le tibia .....</i>	<i>9</i>
1.3. <i>La rotule.....</i>	<i>10</i>
1.4. <i>Le ménisque.....</i>	<i>11</i>
1.5. <i>Le cartilage .....</i>	<i>12</i>
1.6. <i>Les ligaments.....</i>	<i>12</i>
2.    LA GONARTHROSE .....	14
2.1. <i>Définition .....</i>	<i>14</i>
2.2. <i>Les principaux facteurs de l'apparition de la gonarthrose .....</i>	<i>16</i>
2.3. <i>Diagnostic de la gonarthrose.....</i>	<i>20</i>
3.    LA BIOMECHANIQUE DU GENOU .....	22
3.1. <i>Définition de la marche.....</i>	<i>24</i>
3.2. <i>Le cycle de la marche.....</i>	<i>24</i>
4.    LA CINEMATIQUE DU GENOU .....	27
4.1. <i>Les axes du genou .....</i>	<i>27</i>
4.2. <i>Le degré de liberté du genou.....</i>	<i>28</i>
4.3. <i>Les systèmes de mesure du mouvement.....</i>	<i>31</i>
4.4. <i>Les outils de mesure cinématique 3D.....</i>	<i>32</i>
4.5. <i>Le choix du système d'analyse cinématique.....</i>	<i>38</i>
5.    L'ARTHROPLASTIE DU GENOU.....	41
5.1. <i>L'arthroplastie partielle du genou (APG).....</i>	<i>42</i>
5.2. <i>L'arthroplastie totale du genou (ATG) .....</i>	<i>43</i>
5.3. <i>Les fondements de l'intervention de l'arthroplastie du genou.....</i>	<i>44</i>
CONCLUSION.....	44
<b>CHAPITRE 3.....</b>	<b>46</b>
<b>LA CLASSIFICATION DES FORMES .....</b>	<b>46</b>
1.    LA CLASSIFICATION SUPERVISEE .....	48



1.1.	<i>Les k-plus proches voisins</i> .....	48
1.2.	<i>Les arbres de décision</i> .....	49
1.3.	<i>Les réseaux neurones</i> .....	52
1.4.	<i>Naïve bayes</i> .....	53
1.5.	<i>Les machines à support de vecteurs</i> .....	54
2.	LA CLASSIFICATION NON-SUPERVISEE .....	55
2.1.	<i>L'approche du K-moyenne</i> .....	55
2.2.	<i>La classification hiérarchique</i> .....	57
3.	LE BAGGING ET LE BOOSTING .....	59
3.1.	<i>Le bagging</i> .....	60
3.2.	<i>Bootstrap</i> .....	60
3.3.	<i>Agrégation</i> .....	61
4.	LE BOOSTING .....	63
<b>CHAPITRE 4</b> .....		<b>64</b>
<b>CLASSIFICATION DES DONNEES BIOMECHANQUES : METHODOLOGIE ET RESULTATS EXPERIMENTAUX</b> .....		<b>64</b>
1.	BASE DE DONNEES .....	65
2.	LES FORETS ALEATOIRES .....	66
2.1.	<i>Principe et algorithme des forêts aléatoires</i> .....	67
2.2.	<i>L'erreur en dehors du bootstrap ou l'Out-Of-Bag (OOB)</i> .....	69
3.	LES CRITERES D'EVALUATION .....	69
4.	RESULTATS EXPERIMENTAUX ET DISCUSSION .....	71
<b>CONCLUSION GENERALE</b> .....		<b>95</b>
<b>ANNEXES</b> .....		<b>97</b>
5.	ANNEXE A .....	97
6.	ANNEXE B .....	100
7.	ANNEXE C .....	103
<b>BIBLIOGRAPHIE</b> .....		<b>107</b>
<b>WEBOGRAPHIE</b> .....		<b>115</b>

## **LISTE DES TABLEAUX**

Tableau 1 : Nombre d'arthroplasties du genou, par lieu de traitement des patients, 2009-2010 à 2013-2014

Tableau 2 : Répartition des patients ayant subi une arthroplastie du genou, selon le sexe et le groupe d'âge, Canada, 2013-2014

Tableau 3 : Bagging vs Boosting

Tableau 4 : Répartition de la BD en jeux de données

Tableau 5 : Matrice de confusion des données d'abduction-adduction

Tableau 6 : Matrice de confusion des données de flexion-extension

Tableau 7 : Matrice de confusion des données de rotation interne-externe

Tableau 8 : Récapitulatif des taux mesurés par les matrices de confusion

## **LISTE DES FIGURES**

Figure 1 : Anatomie du genou

Figure 2 : Le fémur

Figure 3 : Le tibia

Figure 4 : La rotule (patella)

Figure 5 : Les ménisques vus du haut

Figure 6 : Vue frontale du genou avec le tendon rotulien transparent

Figure 7 : Genou non chirurgical versus genou arthrosique

Figure 8 : Genou arthrosique avec usure du cartilage articulaire

Figure 9 : Genou de l'articulation normale à l'arthrose

Figure 10 : Genou normal et genou avec les vices architecturaux congénitaux

Figure 11 : Gonarthrose sévère de la radiographie de face du genou gauche

Figure 12 : Les trois plans de référence et six directions fondamentales du corps humain par rapport à la position anatomique

Figure 13 : Représentation du cycle de marche d'après Viel

Figure 14 : Représentation du cycle de marche d'après Whittle

Figure 15 : Types de genou

Figure 16 : Les mouvements du genou

[Figure 17 : L'abduction et l'adduction apparaissent dans le plan frontal alors que les rotations interne et externe apparaissent dans le plan transversal](#)

[Figure 18 : Schéma de principe du système magnétique](#)

[Figure 19 : Anatomie du genou vs arthroplastie du genou](#)

[Figure 20 : Arthroplastie partielle du genou \(APG\) ou prothèse de genou unicompartmentaire](#)

[Figure 21 : Arthroplastie totale du genou](#)

[Figure 22 : Arbre hiérarchique constitué de cinq individus a-b-c-d-e et formant trois classes](#)

[Figure 23 : Exemple d'arbre de décision](#)

[Figure 24 : Réseau de neurones](#)

[Figure 25 : Hyper-plan optimal et marge maximale](#)

[Figure 26 : Partitionnement de données avec l'algorithme K-moyennes](#)

[Figure 27 : Illustration d'un tirage aléatoire avec remise pour la formation d'un échantillon](#)

[Figure 28 : Des échantillons bootstrap d'une base de données](#)

[Figure 29 : Fonctionnement de la technique de Bagging](#)

[Figure 30 : Premier arbre de décision des 10 arbres de la forêt des données d'abduction-adduction](#)

[Figure 31 : Description du premier arbre de décision des 10 arbres de la forêt des données d'abduction-adduction](#)

[Figure 32 : Dernier \(10ème\) arbre de décision des 10 arbres de la forêt d'abduction-adduction](#)

[Figure 33 : Description du dernier arbre de décision des 10 arbres de la forêt des données d'abduction-adduction](#)

[Figure 34 : Premier arbre de décision des 10 arbres de la forêt des données de Flexion-extension](#)

[Figure 35 : Description du premier arbre de décision des 10 arbres de la forêt des données de Flexion-extension](#)

[Figure 36 : Dernier arbre de décision des 10 arbres de la forêt des données de Flexion-extension](#)

[Figure 37 : Description du dernier arbre de décision des 10 arbres de la forêt des données de Flexion-extension](#)

[Figure 38 : Premier arbre de décision des 10 arbres de la forêt des données de Rotation](#)

[Figure 39 : Description du premier arbre de décision des 10 arbres de la forêt des données de Rotation](#)

[Figure 40 : Dernier arbre de décision des 10 arbres de la forêt des données de Rotation](#)

Figure 41 : Description du dernier arbre de décision des 10 arbres de la forêt des données de Rotation

À toi ma très chère, et regrettée  
mère, pour l'affection que tu n'as  
jamais cessée de me témoigner,  
je dédie ce mémoire.





## **Avant-propos**

Le présent mémoire est le fruit des efforts consentis tout au long de ma formation au programme de Maîtrise en Informatique. Mes sentiments de gratitude s'adressent d'abord à mon Seigneur, le Tout Puissant qui est Ma Forteresse, Lui qui m'a guidé et inspiré depuis mes premiers pas. Il est Mon Sauveur et Mon berger pour toujours.

Mes sincères et considérables remerciements à ma directrice de mémoire, Dre Neila Mezghani qui a accepté de diriger ce travail, ainsi qu'au Monsieur Youssef Ouakrim pour son soutien et encouragements constants. C'est grâce à leurs conseils techniques, leurs suggestions et connaissances qu'ils m'ont transmis que ce travail a pu être accompli.

Je remercie tous les membres de l'équipe du Laboratoire de recherche en Imagerie et Orthopédie (LIO), pour m'avoir fourni toutes les données nécessaires pour la réalisation de ce travail.

Mes remerciements à tous ceux qui, de près ou de loin, m'ont assisté dans la réalisation de cette œuvre.

Mes remerciements vont également à mes très chers regrettés parents. Enfin je remercie ma famille pour ses sacrifices et son encouragement physique et morale.

De tout cœur, merci à tous!



## Chapitre 1

### Introduction générale

L'intelligence artificielle (IA) est un ensemble de techniques qui visent à s'approcher et imiter le raisonnement humain. Grâce à son efficacité et à la pertinence de ses résultats, l'IA a connu un accroissement de ses domaines d'utilisation et de ses applications. Parmi ses applications nous citons l'aide à la prise de décision qui utilise principalement les techniques de la fouille de données et de reconnaissance de formes. Ces concepts, très performants, sont beaucoup utilisés de nos jours pour l'extraction de connaissances à partir des données en vue d'une prise de décision optimale. Pour ce faire, ils utilisent entre autres des méthodes de classification basées sur des probabilités et la statistique inférentielle. La fouille de données et la reconnaissance de forme présentent une grande importance dans le domaine médical. En effet, ce dernier constitue un excellent champ d'expérimentation pour tester et évaluer les différents paradigmes de l'IA. Pour preuves, plusieurs systèmes dits intelligents ont été conçus afin d'aider la médecine.

Bien que la décision du clinicien soit le facteur le plus déterminant dans un diagnostic, les systèmes d'aide au diagnostic médical gagnent en popularité. Ils sont devenus importants dans beaucoup de disciplines médicales. Ils apportent une aide considérable au clinicien dans le processus de diagnostic où ils réduisent les erreurs dues à l'ambiguïté, au manque d'expériences et à la fatigue par exemple.

Dans un contexte où le Québec est la deuxième province au Canada après l'Ontario où le nombre d'arthroplasties du genou est le plus élevé (Tableau 1), un système décisionnel constituerait un outil adéquat pour assister les cliniciens dans la prise de décision, l'outil serait surtout utile à limiter le nombre de référence inutile vers les chirurgiens, ce qui diminuerait significativement les délais d'attente pour voir le spécialiste. Dans le présent mémoire, nous présentons un travail dont l'objectif consiste à valider si l'arbre de décision corrèle avec la décision du spécialiste à mettre une étiquette de cas chirurgicaux et les non-chirurgicaux d'une arthroplastie. Plus précisément, notre objectif est de développer un système de classification afin de classer un ensemble de sujet candidats pour une arthroplastie en deux classes : chirurgicaux ou non chirurgicaux. Le système sera fondé sur les techniques de forage de données et de reconnaissance de formes et leur application au domaine du diagnostic en milieu clinique.

Lieu de traitement	Arthroplasties du genou					Changement en pourcentage sur 5 ans
	2009-2010	2010-2011	2011-2012	2012-2013	2013-2014	
Terre-Neuve-et-Labrador	676	728	913	925	937	38,6 %
Île-du-Prince-Édouard	235	202	255	316	338	43,8 %
Nouvelle-Écosse	1 645	1 762	1 825	1 898	1 964	19,4 %
Nouveau-Brunswick	1 175	1 224	1 438	1 454	1 465	24,7 %
Québec	7 870	9 132	9 867	10 289	11 207	42,4 %
Ontario	21 768	22 026	23 766	24 964	25 765	18,4 %
Manitoba	1 980	2 058	2 179	2 139	2 268	14,5 %
Saskatchewan	2 192	1 889	1 981	2 465	2 983	36,1 %
Alberta	4 813	5 054	5 870	6 163	6 240	29,6 %
Colombie-Britannique	6 532	6 928	7 155	7 351	6 921	6,0 %
Territoires	60	63	29	58	48	-20,0 %
Canada	48 946	51 066	55 278	58 022	60 136	22,9 %

**Tableau 1 : Nombre d'arthroplasties du genou, par lieu de traitement des patients, 2009-2010 à 2013-2014 (1)**

### **1.1. Contexte, justification et problématique**

En milieu clinique, le médecin généraliste se trouve au quotidien face à des patients pour lesquels, il doit fournir un diagnostic. Dans certains cas, il a aussi besoin d'avoir l'avis d'autres cliniciens, de spécialistes de certains domaines qui ne sont pas toujours disponibles.

Les systèmes experts, premiers systèmes d'aide à la décision, pourraient résoudre ce problème, mais ils sont complexes et nécessitent forcément un expert présent pour valider le modèle avant son utilisation ce qui rend leurs développements moins rapides. D'où l'idée d'utiliser un système qui en lui-même inclut l'expertise de l'expert à partir d'anciennes données. C'est l'objectif du forage de données.

Le forage de données est l'application des techniques de statistiques, d'analyse de données et d'apprentissage automatique à l'exploration d'ensembles de données, en vue d'extraire les connaissances et les informations utiles qu'elles recèlent (Mezghani, 2015). Il permet de faciliter l'extraction de connaissances à partir de grandes quantités de données. Grâce aux données fiables d'anciens patients, le système d'aide au diagnostic basé sur les techniques d'apprentissage automatique peut donner au clinicien, l'information dont il a besoin.

Plusieurs travaux ont été réalisés dans ce domaine afin d'aider les cliniciens dans leurs prises de décisions. A titre d'exemple nous citons, un système de suivi et d'aide à la décision basé sur une ontologie a été conçu dans le domaine de la mammographie (MESSALTI et HEMILA (2013)), un système de suivi basé sur le raisonnement à partir

de cas a été développé en médecine interne (BAKHOUCHE et BENGLIA, 2013), ainsi qu'un système de classification automatique de l'hypothyroïdie (DOUIBI, SETTOUD et AL, 2013).

## **1.2. Objectifs**

L'objectif général du travail est de réaliser un système d'aide à la décision chirurgicale pouvant suggérer si un patient doit subir une arthroplastie du genou ou pas en se basant sur ses caractéristiques cinématiques et en utilisant les forêts aléatoires.

Les objectifs spécifiques de ce projet sont énumérés comme suit :

- Analyser et explorer des données biomécaniques des patients gonarthrosiques. L'accent étant mis sur les données cinématiques;
- Étudier la classification automatique des données cinématiques de deux groupes de sujets : les chirurgicaux et les non chirurgicaux;
- Concevoir et réaliser un système d'aide à la décision pour supporter le besoin d'une arthroplastie basé sur les données cinématiques.

Ce mémoire est organisé en 4 chapitres. Après ce chapitre d'introduction générale, le chapitre 2 présentera le genou et sa biomécanique. Le chapitre 3 détaillera les méthodes de classification de formes identifiées dans la littérature. Dans le chapitre 4,

nous détaillerons le système de classification développé et nous présenterons les résultats expérimentaux. Finalement le chapitre 5 conclura le mémoire.

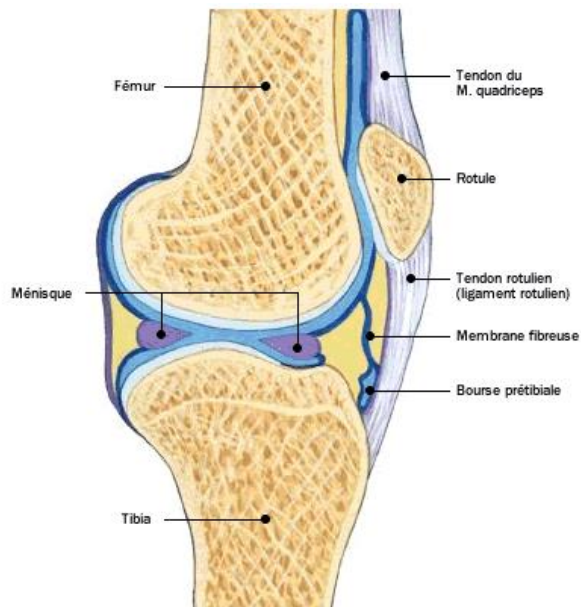


## **Chapitre 2**

### **Le genou et sa biomécanique**

#### **1. L'anatomie du genou**

Le genou est l'une des articulations les plus importantes du système musculo-squelettique. Il est constitué du fémur, le tibia et de la patella ou la rotule qui sont reliés entre eux pour former l'articulation fémoro-patellaire et la double articulation fémoro-tibiale. C'est sur le genou que repose en grande partie le poids du corps lorsque le sujet court, marche ou reste debout (Ouakrim, 2011). La fluidité des mouvements du genou est assurée en outre par le cartilage. Son importance est vitale et il est facile d'imaginer les conséquences que peut occasionner toute lésion. En cas de blessure, les mouvements quotidiens deviennent extrêmement difficiles et douloureux. De nombreuses personnes connaissent malheureusement cette dure réalité. En effet, selon Canadian Institute for Health Information (CIHI), plus de 60.000 patients chirurgicaux du genou ont subi l'arthroplastie entre 2013 et 2014.



**Figure 1 : Anatomie du genou [2]**

### 1.1. Le fémur

Le fémur est l'os de la cuisse. C'est l'os le plus long du corps humain. Il forme au niveau de sa tête, une articulation avec la cavité cotyloïde de l'iliaque et dans une autre partie, il forme l'articulation du genou avec le tibia, le péroné et la rotule au niveau du condyle. Tout le poids du corps se repose directement sur le fémur (Figure 2).

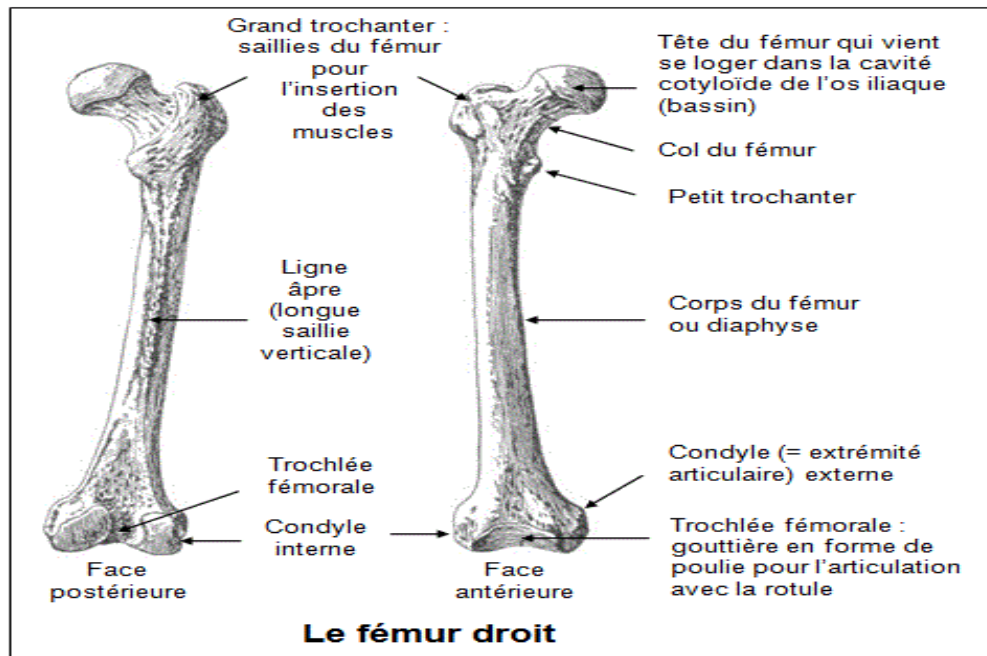
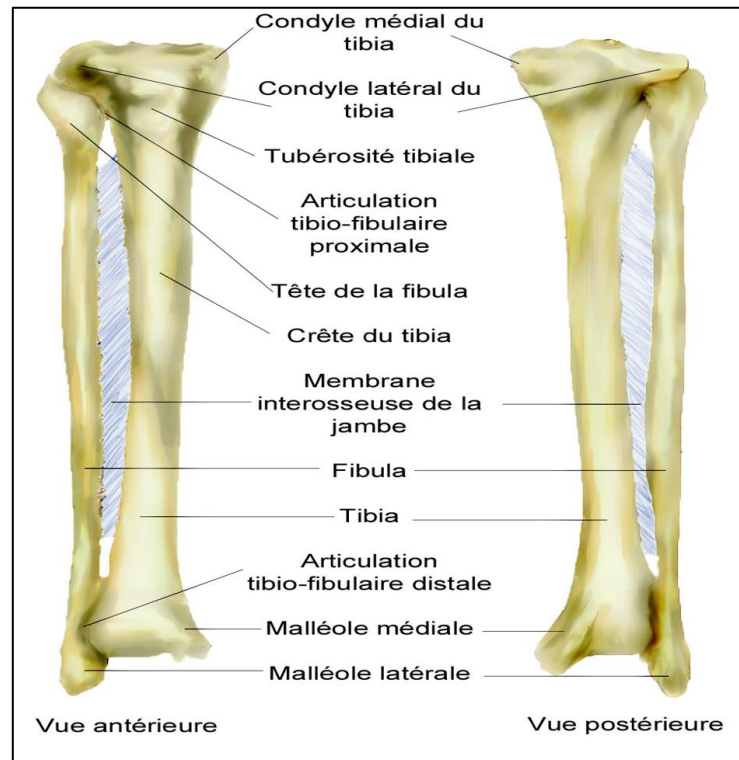


Figure 2 : Le fémur [3]

## 1.2. Le tibia

Le corps humain se repose essentiellement sur le tibia. Il forme à son extrémité supérieure, l'articulation du genou avec le péroné et le fémur tandis qu'il forme l'articulation de la cheville à son extrémité inférieure, avec le péroné et l'astragale du pied. La partie inférieure du tibia est terminée par la malléole qui constitue la cheville (Figure 3).

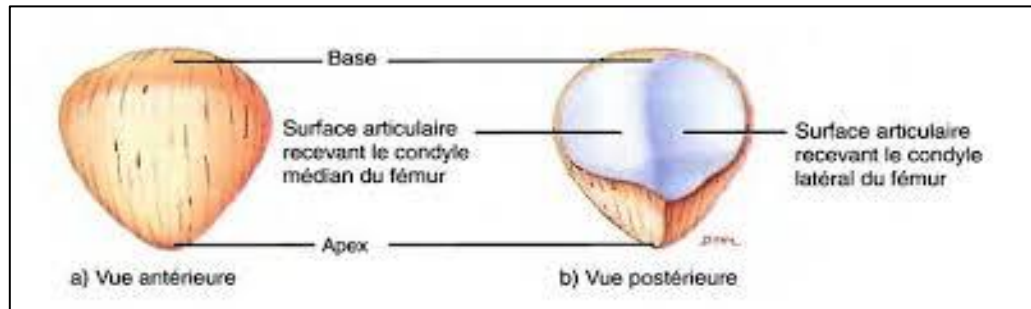
Le fémur et le tibia permettent à l'individu de faire sur le plan sagittal, le mouvement de la flexion-extension, sur le plan transversal, le mouvement de rotation et sur le plan frontal, les mouvements d'abduction et d'adduction.



**Figure 3 : Le tibia [4]**

### 1.3. La rotule

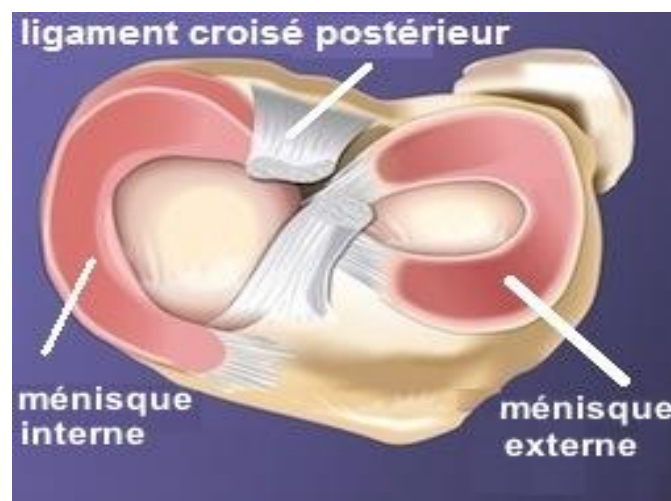
La rotule ou la patella est le petit os qui s'articule avec le tibia et le fémur. Elle est maintenue par deux tendons : le tendon du quadriceps et rotulien. La rotule est comme un bras de levier dans l'extension du genou (Figure 4).



**Figure 4 : La rotule (patella) [5]**

#### 1.4. Le ménisque

Le ménisque est une structure cartilagineuse en forme « C » qui est fixé au tibia et qui agit comme un amortisseur pour le genou. La déchirure du ménisque se traduit par une douleur dans le genou et ainsi qu'un blocage du genou en flexion avec impossibilité d'étendre la jambe.



**Figure 5 : Les ménisques vus du haut [6]**

### **1.5. Le cartilage**

Le cartilage est un tissu constitué de cellules rondes qui protègent les extrémités du fémur et du tibia afin de leur faciliter les mouvements. Le cartilage a une double fonction: une fonction de protection contre les chocs biomécaniques et une autre qui permet aux surfaces articulaires de glisser de façon harmonieuse à cause de ses caractéristiques d'élasticité (Ouakrim, 2011).

### **1.6. Les ligaments**

Les ligaments sont des structures robustes qui stabilisent l'articulation de genou. On distingue les ligaments croisés et les ligaments collatéraux.

#### *1.6.1. Les ligaments croisés*

Les ligaments croisés se croisent de l'avant vers l'arrière dans l'articulation du genou. Ils se situent à l'intérieur du genou. Ils forment la lettre « X » et stabilisent l'articulation en limitant la rotation interne et en réduisant le déplacement antéro-postérieur du tibia lors de la flexion.

### 1.6.2. Les ligaments collatéraux

Situés à l'extérieur de la capsule du genou, les ligaments collatéraux stabilisent le genou pour limiter le stress en valgus et en varus autrement dit les sollicitations excessives dans le plan frontal en le fémur et le tibia.

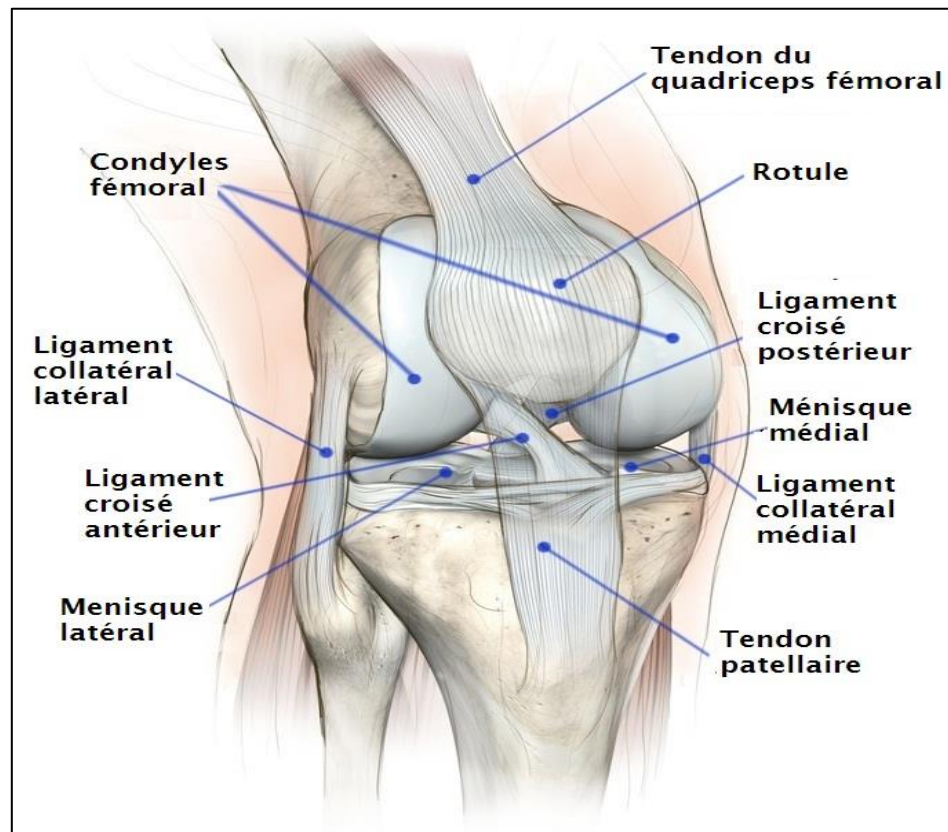


Figure 6 : Vue frontale du genou avec le tendon patellaire transparent [7]

## 2. La gonarthrose

### 2.1. Définition

L'arthrose du genou ou gonarthrose, est une maladie dégénérative qui s'attaque au cartilage des articulations. Selon Canadian Institute for Health Information (CIHI) il touche plus de 80% des personnes âgées de plus de 60 ans (Tableau 2) ainsi que certains sportifs. Quand le cartilage dégénère les articulations souffrent, se déforment et génèrent des malaises. C'est une maladie qui est beaucoup plus fréquente à la vieillesse et chez les personnes à surcharge de poids ou obèses.

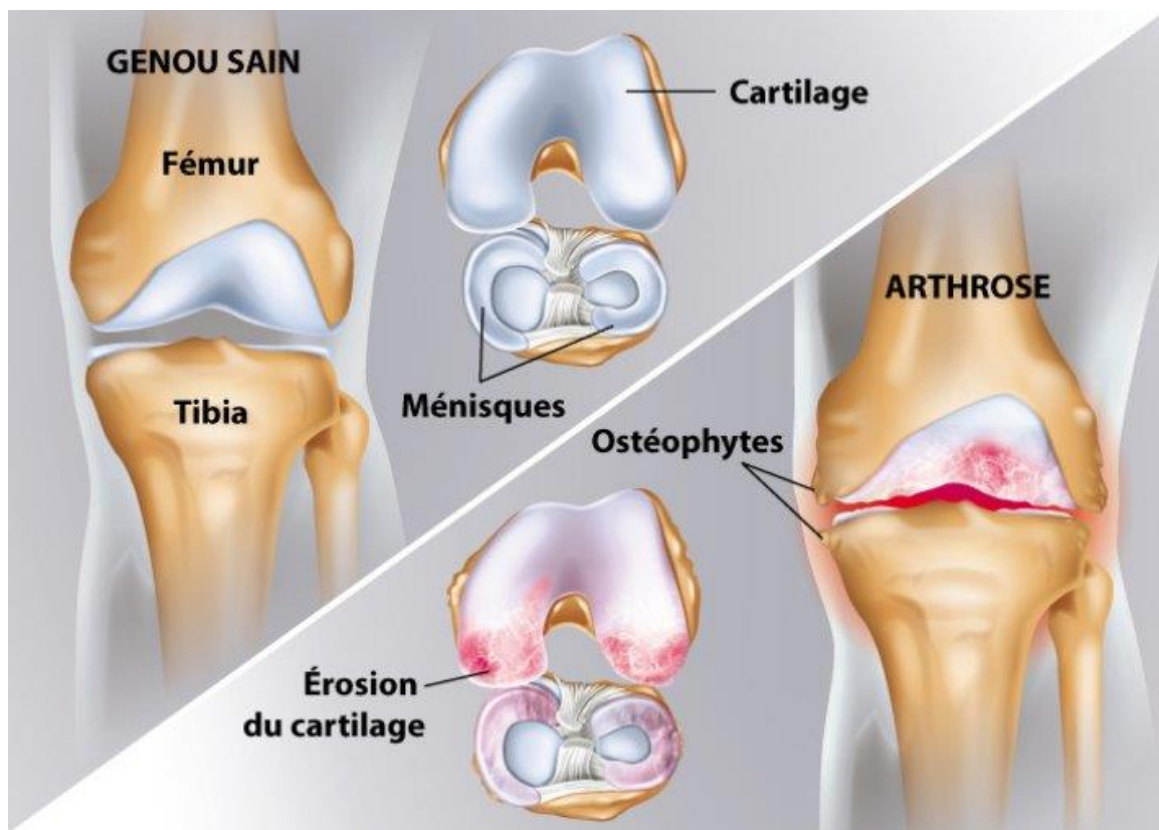
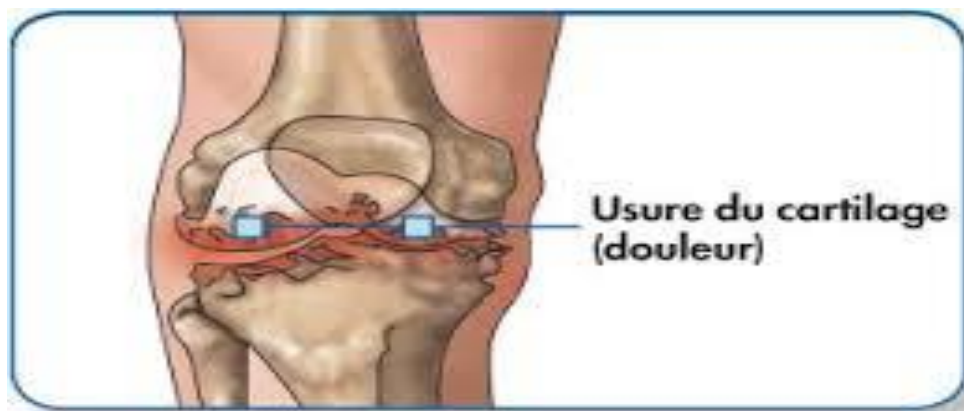


Figure 7 : Genou non chirurgical versus genou arthrosique [8]



Dans le cas normal, appelé aussi non pathologique ou asymptomatique, le cartilage diminue le frottement entre les différentes parties osseuses. La dégradation du cartilage articulaire affecte la bonne répartition des forces sur l'articulation, ce qui nuit aux propriétés biomécaniques et provoque des douleurs chroniques dans l'articulation du genou (Lequesne et Menkes, 1995). La gonarthrose est observée lorsqu'il y a un déséquilibre entre la résistance des tissus articulaires et l'articulation du genou ou, encore, lorsqu'il y a usure du cartilage articulaire qui couvre les extrémités osseuses de l'articulation.



**Figure 8 : Genou arthrosique avec usure du cartilage articulaire [9]**

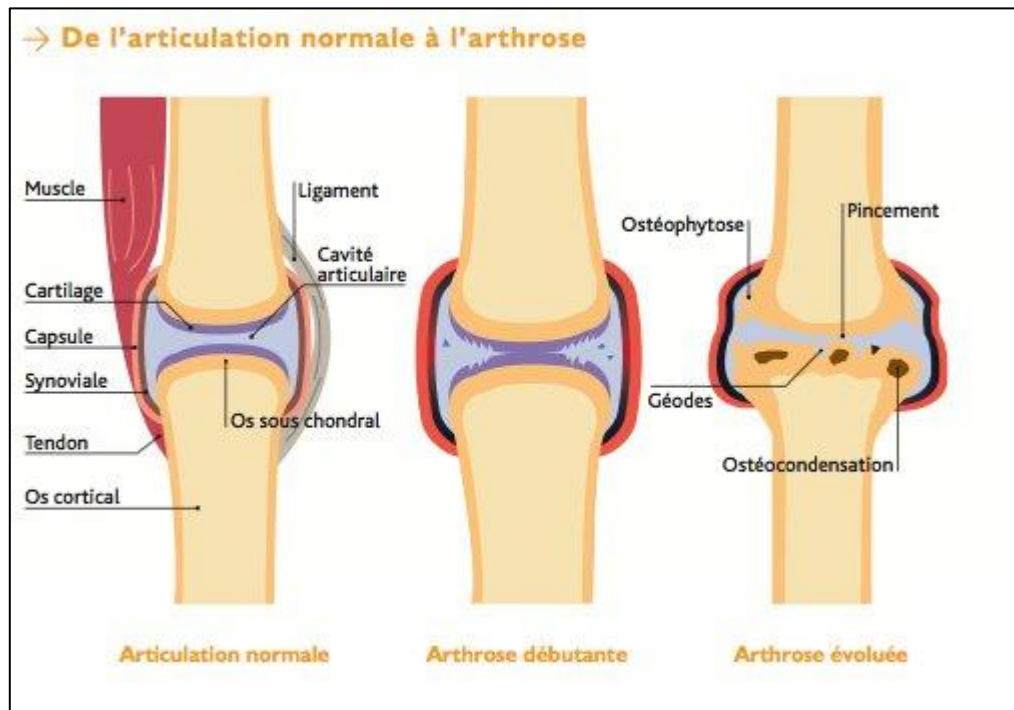


Figure 9 : Genou de l'articulation normale à l'arthrose [10]

## 2.2. Les principaux facteurs de l'apparition de la gonarthrose

La gonarthrose n'est pas qu'une usure de cartilage, elle est aussi une maladie chronique multifactorielle. Cela veut dire qu'il y a plusieurs facteurs qui peuvent être à l'origine de l'arthrose du genou: génétiques, métaboliques, immobilismes et certains facteurs mécaniques (Ouakrim, 2011). Elle est donc la maladie de toute l'articulation.

Les principaux facteurs de risques contribuant à l'apparition de la gonarthrose ont fait l'objet de plusieurs études (Felson, 1995; Hunter, 2009). Ils peuvent être classés en facteurs modifiables et facteurs non modifiables. Ainsi donc, les facteurs tels que

l'âge, le sexe, la génétique et blessures et antécédents de blessures (déchirure ligamentaire, déchirure méniscale, fracture, subluxation, etc.) sont non modifiables tandis que les facteurs tels que l'obésité, le stress sur l'articulation causé par un mouvement anormal du genou ou une charge excessive, un mauvais alignement des jambes, la quantité d'une activité physique qui sollicite l'articulation, l'hypermobilité de l'articulation et les troubles musculaires peuvent être considérés comme modifiables (Cicuttini, Spector et Baker, 1997). Ces principaux facteurs peuvent être considérés comme la base de l'apparition de la pathologie.

En 1986, l'American College of Rheumatology (ACR) a ciblé 3 à 4 critères non-radiologiques dont l'âge supérieur à 50 ans comme un critère clinique de présenter une gonarthrose (Altman et al., 1986). L'Institut Canadien d'Information sur la Santé (ICIS) a montré que la prévalence de la maladie est d'environ de 9 % dans la tranche d'âge – 45 à 54 ans – et de 67% dans la tranche – 55 à 74 ans. Lawrence et al., (2008) qui ont confirmé l'augmentation de la prévalence de la gonarthrose avec l'âge. De même, l'étude de Decade (2005) a démontré une relation directe et exponentielle entre l'âge et la prévalence de la gonarthrose. Les recherches menées par Arden et Nevitt (2006) classent la gonarthrose en deuxième position dans la liste des pathologies qui causent l'incapacité fonctionnelle chez les personnes de 50 ans et plus. L'arthrose reprend une pathologie engendrant d'énorme coût au système de santé.

Une revue de la littérature a démontré que les femmes ont un risque plus élevé d'être atteintes de l'arthrose du genou et de la main quel que soit leur âge (Srikanth et

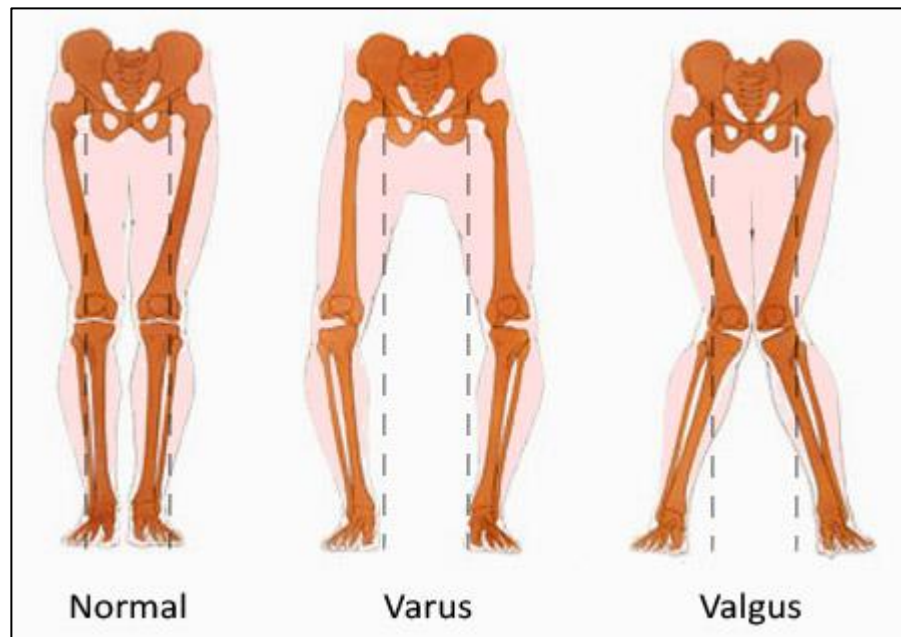
al., 2005). Le risque est encore plus élevé chez les personnes âgées de 50 à 55 ans (Lawrence et al., 2008).

La gonarthrose a tendance à se manifester davantage chez les personnes obèses que chez les personnes non obèses (Cicuttini, Spector et Baker, 1997). Le risque de la gonarthrose est majoré de 15 % pour chaque augmentation d'une unité d'indice de masse corporelle (IMC) (Anderson et Felson, 1988). Le risque de souffrir d'une gonarthrose augmente de 35 % pour chaque augmentation du poids corporel de cinq kilogrammes (Hart et Spector, 1993).

Groupe d'âge	Femmes		Hommes		Femmes et hommes	
	N	%	N	%	N	%
Moins de 45 ans	293	0,8 %	147	0,6 %	440	0,7 %
45-54 ans	3 076	8,5 %	1 821	7,6 %	4 897	8,1 %
55-64 ans	10 871	30,0 %	7 301	30,5 %	18 172	30,2 %
65-74 ans	13 266	36,6 %	8 901	37,2 %	22 167	36,9 %
75-84 ans	7 597	21,0 %	5 114	21,4 %	12 711	21,1 %
85 ans et plus	1 107	3,1 %	642	2,7 %	1 749	2,9 %
Total	36 210		23 926		60 136	

**Tableau 2 : Répartition des patients ayant subi une arthroplastie du genou, selon le sexe et le groupe d'âge, Canada, 2013-2014 [\[1\]](#)**

Certains chercheurs pensent que la pathologie est héréditaire et peut se transmettre de génération en génération de même un mauvais alignement des membres inférieurs (Lawrence et al., 2008) et peut causer des troubles importants du fonctionnement et de la biomécanique du genou. Ces défauts ou « vices architecturaux congénitaux » peuvent s'observer fréquemment chez les hommes qui ont le genou varum (les jambes prennent la forme de parenthèses) et chez les femmes qui ont le genou valgum (Les jambes prennent la forme de X).



**Figure 10 : Genou normal et genou avec les vices architecturaux congénitaux [11]**

Outre les malformations de naissance, le vieillissement et le surpoids jouent un rôle important dans le développement de cette maladie. Quant au surpoids, il augmente les contraintes mécaniques sur l'articulation.

### **2.3. Diagnostic de la gonarthrose**

Le diagnostic de l'arthrose passe essentiellement par un examen clinique. Habituellement, le clinicien peut diagnostiquer l'arthrose après avoir questionné le patient sur les symptômes (nature et intensité de ses douleurs, présence d'une gêne

articulaire) et après avoir fait soigneusement un examen clinique. Il va notamment déterminer si la douleur se déclenche à la pression ou en faisant bouger l'articulation et l'importance de la raideur articulaire. Une articulation enflée est aussi un signe pouvant conduire au diagnostic de l'arthrose. Généralement, les douleurs arthrosiques sont déclenchées par l'effort, mais se calment au repos. Cependant, certaines arthroses s'accompagnent de poussées inflammatoires douloureuses réveillant le patient en pleine nuit et générant un dérouillage matinal.

Par la suite, le clinicien peut demander une radiographie pour l'aider à confirmer son diagnostic et constater le degré de la dégénérescence de l'articulation (légère à sévère ou de grade I à IV – c'est l'échelle de Kelgren Lawrence). Elle ne reflète toutefois pas la condition clinique (symptômes, limitations) et ne permet pas ultimement d'intervenir sur les facteurs de risque modifiables de l'arthrose (Ouakrim, 2011). Les guides de pratique Clinique indiquent que l'IRM est adéquate que lorsqu'il y a un blocage de l'articulation et donc pour investiguer l'état des ménisques. Le recours à une imagerie par résonance magnétique (IRM) s'avère nécessaire afin de caractériser les structures articulaires ou vérifier si l'usure est localisée à un endroit précis.

La radiographie se doit d'être prise alors que les genoux sont en charge (position debout) et doit comporter en outre une vue tangentielle de la rotule pour bien caractériser la sévérité de l'arthrose et permettre de faire un suivi structural de la pathologie (Buckland-Wright, 2006).

Dans la majorité des cas, la radiographie est le seul examen nécessaire pour confirmer le diagnostic de la maladie. Toutefois, très rarement, d'autres examens sont demandés lorsqu'il demeure un doute sur le diagnostic. Il peut alors s'agir d'un scanner, d'une scintigraphie osseuse, d'une arthrographie, d'une échographie ou encore d'une IRM.

La biopsie synoviale n'est vraiment employée qu'à titre exceptionnel, son intérêt pour le diagnostic étant très restreint.

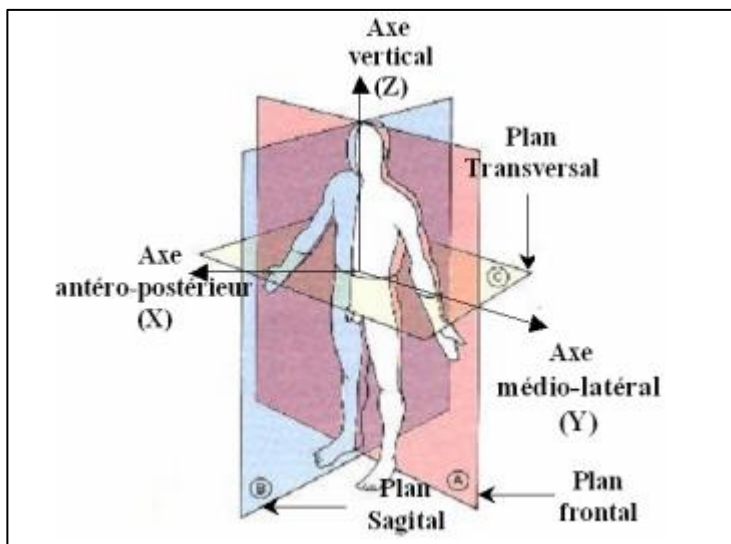
### **3. La biomécanique du genou**

L'analyse de la mécanique de la marche humaine nécessite d'introduire certaines terminologies basiques d'anatomie. Le mouvement des segments corporels est décrit comme apparaissant dans trois plans de l'espace tridimensionnel qui sont référencés par rapport à la position anatomique de l'individu ([figure 12](#)). La position anatomique est la position dans laquelle un individu est debout, droit, les pieds parallèles, les bras de chaque côté du corps et les paumes de main dirigées vers l'avant. Le plan sagittal divise le corps en portions droite et gauche. Le plan frontal divise le corps en portions antérieure et postérieure. Le plan frontal est aussi connu comme le plan coronal. Le plan transverse divise le corps en portions supérieure et inférieure. A l'intérieur d'un segment corporel, des positions anatomiques relatives sont décrites en utilisant des termes spécifiques. La position médiale définit la position sur la ligne médiane du segment. La position latérale décrit une position en dehors de cette ligne. La position proximale représente la position anatomique la plus proche du point de référence en



général, le centre de masse du sujet. La position distale définit la position la plus éloignée du point de référence. Les déplacements articulaires au cours de la marche peuvent être décrits en utilisant les définitions du mouvement dans les trois plans de référence. Les mouvements du genou dans les plans de référence sont illustrés sur la [figure 13](#).

En biomécanique, les mouvements sont généralement décrits dans un référentiel. Ce dernier comporte trois plans et trois axes décrits sur la Figure 13.



**Figure 12 : Les trois plans de référence et six directions fondamentales du corps humain par rapport à la position anatomique (Williams, 1986).**

### 3.1. Définition de la marche

La marche humaine normale peut être définie comme un phénomène complexe qui implique la coordination de mouvements de rotations des segments corporels pour maintenir l'équilibre du corps pendant son déplacement vers l'avant (Perry, 1992). Elle s'acquiert (Inman et al., 1981) à travers un processus d'apprentissage et les patrons (en anglais *patterns*) caractéristiques de la marche dite normale, ont été appréhendés afin d'être identifiés.

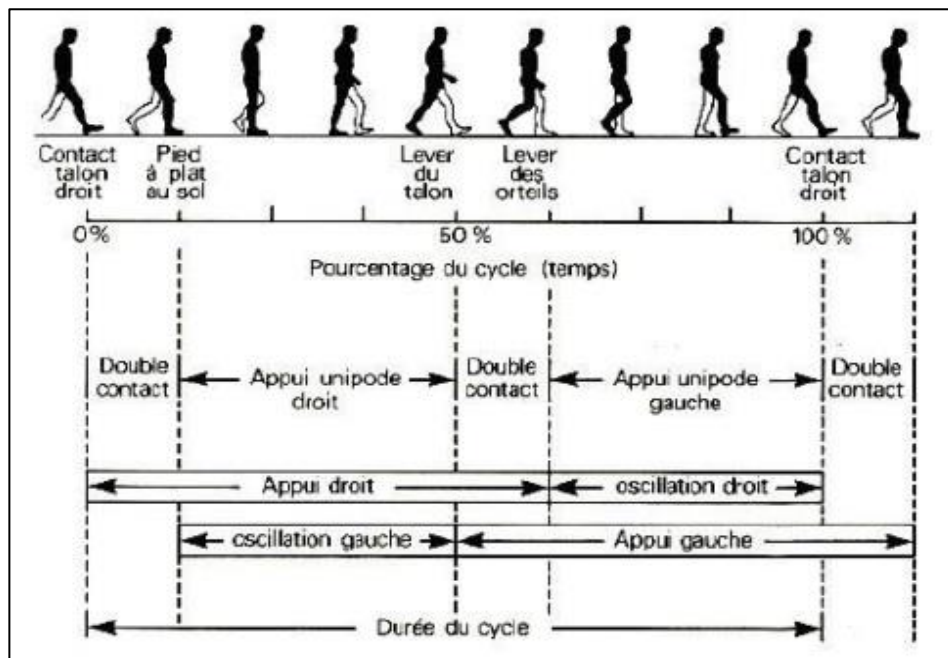
### 3.2. Le cycle de la marche

L'objectif premier de la marche humaine est de déplacer le centre de masse (CM) dans la direction de progression du mouvement. Le plan de progression est parallèle au plan sagittal. L'une des caractéristiques principales de la marche est qu'elle s'accomplit suivant un patron de mouvements cycliques qui apparaît à chaque pas. Un cycle de marche découpé en diverses phases sert de référentiel de discussion à l'analyse de la marche humaine. Les résultats de ces analyses sont en général normalisés et les différentes phases sont exprimées en pourcentage de celui-ci. La [figure 14](#) illustre un cycle de marche. Celui-ci regroupe la phase d'appui qui représente environ 60% et d'oscillation qui représente environ 40% du cycle. Le cycle de marche peut également être découpé en phases de simples et doubles appuis. La phase de simple appui (ou phase d'oscillation du membre controlatéral) est définie quand le pied est en contact avec le sol pendant que le pied du membre controlatéral oscille. Elle commence à

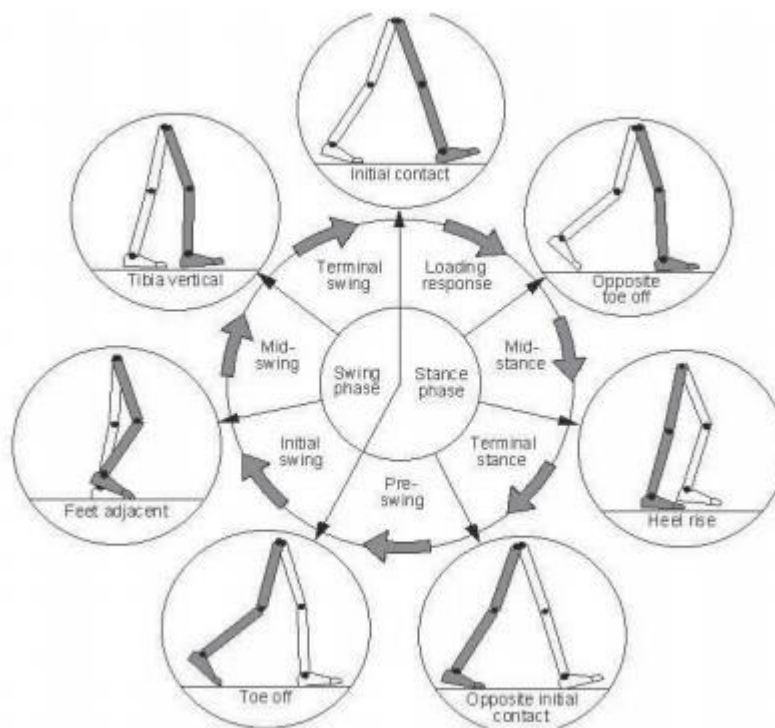
l'instant de lever des orteils (LO) du pied oscillant et se termine à l'instant de contact initial (CI) du pied oscillant au sol. Pendant la phase de double appui, les deux pieds sont au sol pendant que l'ensemble du corps se déplace vers l'avant. Au cours du double appui, le poids du corps est transféré du pied arrière vers celui de l'avant. Dans la littérature, il est courant de trouver un découpage plus fin du cycle de marche tel que décrit à la [figure 14](#) :

- La phase de chargement du cycle de marche commence à CI (0 à 2%) et se termine environ à LO. Au cours de la phase de freinage, le CM est transféré vers la jambe avant en appui au sol. Elle permet également d'absorber le choc dû au contact du pied au sol en gardant une certaine stabilité.
- Le début d'appui. Il s'agit de la première moitié de la phase de simple appui, permettant au corps de se déplacer vers l'avant au-dessus du membre inférieur en appui. Elle prend fin à l'instant où le CM atteint son apogée (e.g. midstance).
- La fin d'appui représente entre 30 et 50%. C'est la seconde moitié de la phase de simple appui, représentant le déplacement du corps vers l'avant jusqu'au contact initial du pied controlatéral (CIC) au sol.
- La phase de propulsion représente entre 50 et 60%. Elle correspond globalement à la deuxième phase de double appui.
- L'oscillation débute entre 60 et 73%.

– La fin d'oscillation correspond à l'intervalle 86 à 100%. La jambe qui réalise la fonction de support est appelé jambe d'appui. Lors de la phase de double appui, puisque les deux membres inférieurs sont en contact avec le sol, la jambe d'appui est la jambe avant puisque celle-ci réalise la fonction de support. La distance parcourue vers l'avant par le pied oscillant de LO à CI est définie comme la longueur de pas. La cadence est définie comme le nombre de pas par unité de temps (en pas.min<sup>-1</sup>).



**Figure 13 : Représentation du cycle de marche d'après Viel (Viel, 2000).**



**Figure 14 : Représentation du cycle de marche d'après Whittle (Whittle, 2002).**

#### **4. La cinématique du genou**

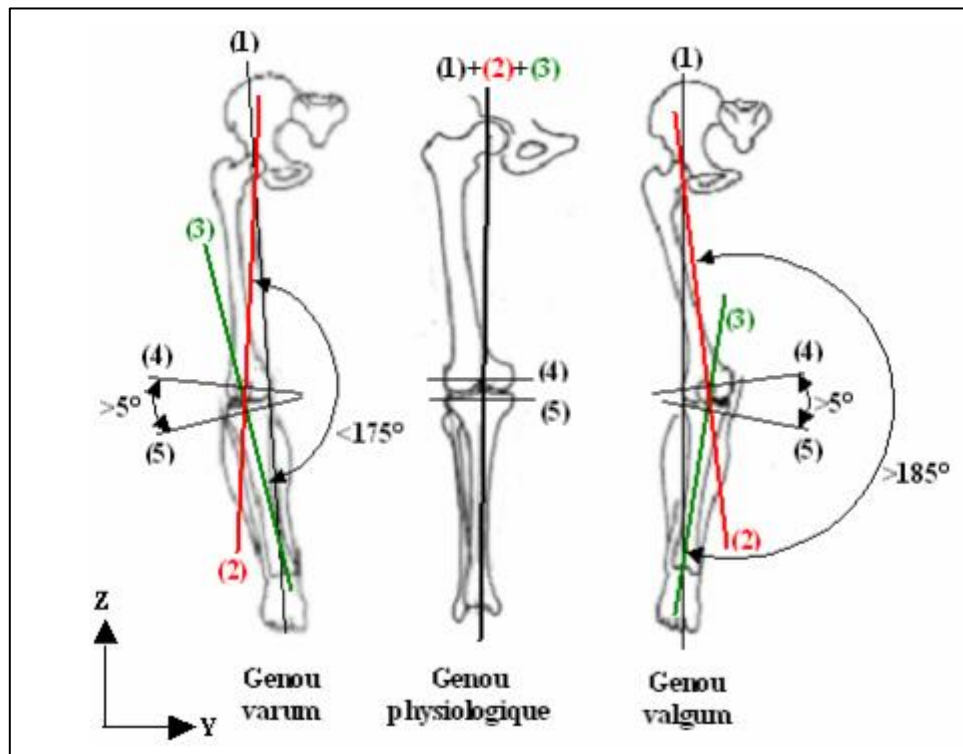
##### **4.1. Les axes du genou**

L'axe mécanique du genou passe par le milieu de la tête du fémur, le centre de l'articulation fémoro-tibiale et le milieu de la poulie astragaliennne.

#### 4.2. Le degré de liberté du genou

On peut attribuer six degrés de liberté à l'articulation du genou. Les trois premières libertés s'observent lors des mouvements de flexion et extension du genou dans le plan sagittal et les trois derniers pendant les mouvements de rotation dans le plan frontal et le plan transversal.

C'est ainsi qu'on peut alors mesurer l'angle entre les milieux des articulations de la hanche, du genou et de la cheville dans le plan frontal et déterminer les types de genou (varum, physiologique, valgum).



**Figure 15 : Types de genou (d'après Kapandji, 1996) ; (1) : axe joignant le centre de la tête du fémur et celui de la cheville (axe mécanique du membre inférieur), (2) : axe mécanique du fémur, (3) : axe mécanique du tibia, (4) : axe des condyles, (5) : axe des plateaux tibiaux.**

#### *4.2.1. Les mouvements de flexion et d'extension*

Les déplacements dans le plan sagittal entraînent les mouvements de flexion et d'extension. Ce sont des mouvements de roulement et de glissement des condyles fémoraux sur le tibia. Au début, la flexion commence par un glissement isolé, ensuite un roulement s'observe lentement jusqu'à l'atteinte de 60% du mouvement de flexion et extension (entre 60° et 90°) et, finalement, le roulement diminue et la flexion se limite au glissement pur (Ouakrim, 2011).

#### *4.2.2. Les mouvements de rotation*

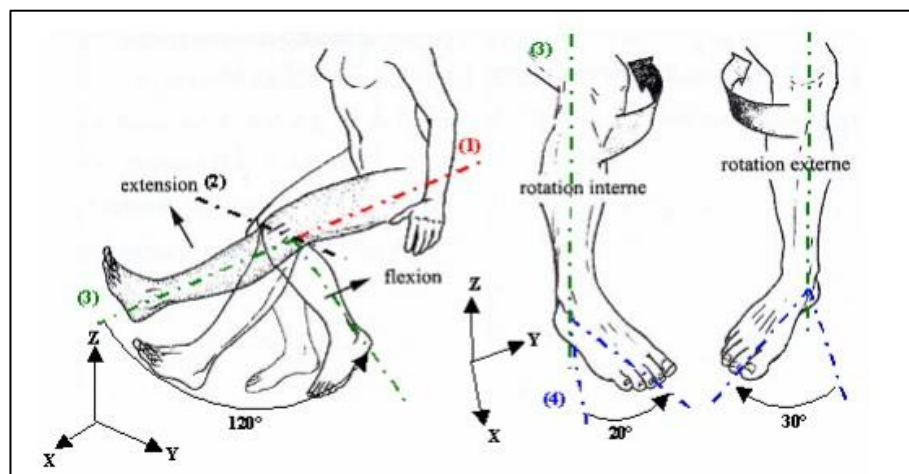
Les mouvements susceptibles d'apparaître dans le plan transversal sont les rotations interne et externe. Quand le genou est fléchi, la partie postérieure des condyles est en contact avec la partie moyenne des glènes. Le tibia est considéré fixe.

En rotation interne, le condyle médial pivote et glisse vers l'avant ; le condyle latéral pivote et glisse vers l'arrière.

En rotation externe, le condyle latéral avance puis glisse, le condyle médial recule puis glisse.

Les ménisques suivent les mouvements des condyles sur les glènes. C'est à dire qu'en rotation interne, le ménisque externe recule et le ménisque interne avance alors qu'en rotation externe, le ménisque externe avance, le ménisque interne recule.

En rotation interne, la patella regarde vers l'intérieur tandis qu'en rotation externe elle est tournée vers le dehors.



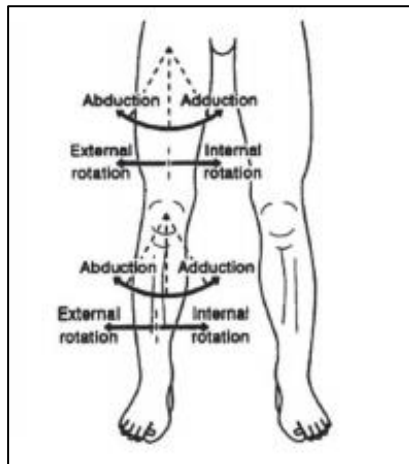
**Figure 16 : Les mouvements du genou (d'après Calais-Germain, 1986) ; (1) : axe longitudinal du fémur, (2) : axe des condyles, (3) : axe longitudinal du tibia, (4) : axe longitudinal du pied.**

#### *4.2.3. L'abduction et l'adduction*

Les mouvements dans le plan frontal sont l'abduction et l'adduction. C'est le mouvement du tibia versus le fémur en varus-valgus. Le mouvement d'adduction-



abduction est calculé par les mouvements angulaires sur le plan frontal perpendiculaire à l'axe médio-latéral du fémur et à l'axe longitudinal du tibia (Ouakrim, 2011).



**Figure 17 : L'abduction et l'adduction apparaissent dans le plan frontal alors que les rotations interne et externe apparaissent dans le plan transversal (Whittle, 2002).**

#### **4.3. Les systèmes de mesure du mouvement**

Les systèmes de mesure du mouvement peuvent être divisés en trois catégories : les systèmes de mesure cinématique, les instruments de mesure dynamique et l'électromyographie (Gorce, 2000). L'électromyographie (EMG) est une méthode d'enregistrement de l'activité électrique musculaire à l'aide d'électrodes placées sur les muscles à étudier. Il existe des électrodes de surface qui enregistrent l'activité musculaire à travers la peau, et des électrodes intramusculaires pour l'étude des

muscles profonds. Nous ne développerons pas cette technique car nos travaux n'ont pas pour objectif de quantifier l'activité musculaire. Dans cette étude, seuls les principaux moyens de mesure cinématique et de mesure dynamique seront développés.

#### **4.4. Les outils de mesure cinématique 3D**

En biomécanique, l'analyse cinématique consiste à décrire les variations de position des segments corporels dans l'espace et dans le temps. Les méthodes de mesure cinématique tridimensionnelle de la marche utilisent des technologies à base de capteurs (goniomètres, accéléromètres, gyroscopes magnétiques) ou à base de vision (caméras, systèmes optoélectroniques) (El Hafi, 2000). Les méthodes couramment utilisées dans le domaine de la biomécanique sont les goniomètres articulaires (Chao, 1980; Gaitosik, 1987), les accéléromètres (Padgaonkar, 1975), les capteurs de positions magnétiques (Milne, 1996) et les systèmes optoélectroniques (Slobounov, 1999).

##### *4.4.1. Les techniques à base de capteur*

#### **Les goniomètres**

Les goniomètres ont souvent été utilisés pour analyser le cycle de marche. Ils sont composés de deux branches fixées à un potentiomètre (capteurs angulaires). L'information mesurée est une différence de potentiel qui traduit une variation d'angle

entre deux segments corporels. Chaque branche est attachée à un segment et l'axe du goniomètre se place sur l'axe de rotation théorique de l'articulation. La pose et l'étalonnage des goniomètres sont très importants pour obtenir une bonne reproductibilité et fiabilité des mesures (Faivre, 2003). Les potentiomètres à résistances variables donnent une valeur analogique de l'angle d'ouverture et les codeurs optiques incrémentaux fournissent des valeurs analogiques des positions angulaires.

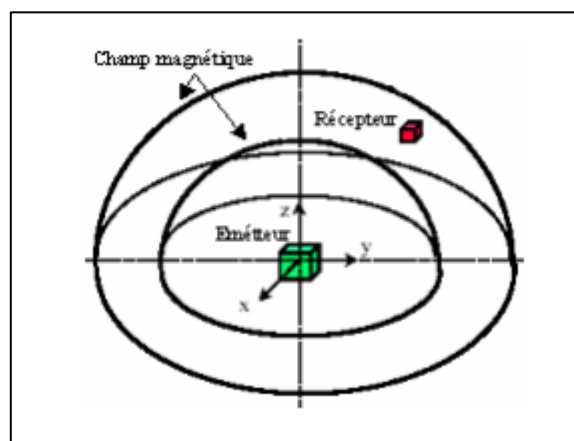
### **Les accéléromètres**

Les accéléromètres sont des capteurs mesurant l'accélération subie par un solide en mouvement. Ils utilisent le principe de la dynamique liant la force à l'accélération d'une masse. L'objectif est de quantifier les forces d'inertie qui s'appliquent sur la masse en mouvement. On mesure à la sortie des accéléromètres une tension ou une résistance proportionnelle à la force appliquée. La masse étant connue et constante, l'accélération peut être calculée. Pour connaître la position d'un point du corps en mouvement, deux intégrations de l'accélération sont nécessaires ainsi que la position initiale du point de départ du mouvement.

### **Les capteurs de positions magnétiques**

Cette technique de capture du mouvement utilise les propriétés des champs magnétiques. Le système est formé de deux ensembles d'antennes orthogonales entre

elles. Le premier ensemble est placé à l'intérieur d'un émetteur (générateur électromagnétique) et le second à l'intérieur d'un récepteur (capteur). Le signal (champ magnétique) reçu par le récepteur permet de déduire la position et l'orientation de celui-ci par rapport à l'émetteur (An, 1988) (Figure 18).



**Figure 18 : Schéma de principe du système magnétique (d'après An, 1988)**

### **Les systèmes optoélectroniques**

Les informations sur les systèmes optoélectroniques ont été recueillies dans les ouvrages d'Allard et Blanchi (1996), Whittle (1996) et Viel (2000). La plupart des systèmes optoélectroniques d'analyse gestuelle utilisent la méthode de stéréovision. Celle-ci permet de déterminer la position spatiale d'un objet à partir de la mise en correspondance de deux images planes. Cette technique est utilisée pour la reconstruction tridimensionnelle des positions de marqueurs à partir d'images

provenant de caméras (Chèze, 1993). La lumière infrarouge est souvent utilisée pour la détection des marqueurs dans les images des caméras. Le rayon de lumière passe par le centre de la focale de l'objectif (centre optique) et vient exciter un ensemble de pixels sur la matrice de capteurs de la caméra. Cet ensemble de pixels se matérialise en une tache lumineuse dont on calcule le barycentre. Connaissant l'emplacement de la focale et les coordonnées de la position du barycentre de la tache lumineuse sur la matrice de la caméra, on calcule les coordonnées de la droite de projection du marqueur. Ainsi, les informations des droites de projection issues d'au moins deux caméras permettent grâce à un algorithme de mise en correspondance de calculer les coordonnées tridimensionnelles du marqueur. La position des caméras dans l'espace doit être connue pour que l'algorithme puisse fonctionner, une phase de calibrage est donc nécessaire avant toute acquisition.

Classiquement, les systèmes optoélectroniques utilisent deux techniques, la première utilise des marqueurs actifs, et la deuxième emploie des marqueurs passifs.

### **Les systèmes à marqueurs actifs**

La lumière est émise par le marqueur lui-même, appelé LED (Light Emitting Diode). Les avantages des systèmes à marqueurs actifs résident dans l'identification automatique des marqueurs et une précision remarquable des trajectoires spatiales, de l'ordre du dixième de millimètre sur les coordonnées tridimensionnelles (Viel, 2000). Par ailleurs, les systèmes les plus récents peuvent fournir en temps réel les trajectoires.

Par contre, ces systèmes sont limités en nombre de capteurs à cause du compromis entre la fréquence de multiplexage et le temps de traitement des signaux. De plus, les LEDs doivent être alimentées par des fils, ce qui non seulement procure une gêne mais aussi limite les distances dans l'exécution de mouvements. Les systèmes Selspot, Optotrack, Polaris et Watsmart utilisent la technique à base de marqueurs actifs. Woltring et al. (1980) testent le système Selspot et concluent qu'il est efficace pour l'étude clinique de la marche en temps réel. Brenière et al. (1991) utilisent le même système pour étudier la phase d'initiation de la marche.

### **Les systèmes à marqueurs passifs**

Un projecteur à impulsion composé de LEDs est fixé sur les caméras autour de leur objectif et le marqueur est recouvert d'un scotch light ayant la particularité de renvoyer le rayon incident dans la même direction (mire rétro-réfléchissante). Les marqueurs n'ayant pas besoin de fils d'alimentation, aucune contrainte n'est imposée pour l'exécution du mouvement. Ce dispositif permet l'enregistrement de gestes complexes (études de sauts périlleux par exemple). Le champ d'acquisition peut être beaucoup plus large selon les types et nombre de caméras utilisés contrairement aux systèmes à base de capteurs actifs. De plus, le nombre de marqueurs n'est pas limité. Cependant, cette technique requiert une phase dite de « dépouillement » pendant laquelle l'opérateur identifie chaque marqueur. Actuellement, la plupart des systèmes intègrent une routine informatique d'auto-labélisation. Ainsi, l'opérateur identifie l'ensemble

des marqueurs sur une image de référence et le reste est identifié automatiquement. La précision des trajectoires tridimensionnelles dépend du champ d'acquisition. Pour plusieurs mètres carrés, elle est de l'ordre de quelques millimètres sur les trajectoires spatiales (El hafi, 1999; Viel, 2000). Parmi les systèmes existants, nous pouvons citer : Elite, Vicon, Ariel, Saga3, Motion Analysis et Smart. De très nombreuses études d'analyse cinématique de la marche sont réalisées à l'aide de ces systèmes dans des domaines d'applications très variés. Les cliniques s'équipent de plus en plus en système optoélectronique pour l'étude de la rééducation fonctionnelle de la marche (Wittle, 1996; Viel, 2000) liée soit à des troubles moteurs comme l'hémiplégie (Pélissier, 1997; Watelain, 1999), soit à la suite d'une chirurgie comme : la pose de prothèses articulaires, les amputations (Goujon, 2006) et les ligamentoplasties (Barbier, 1994). Les laboratoires en informatique et mécanique utilisent ces systèmes pour la simulation virtuelle (Boulic, 1990; Multon, 1998) ou le contrôle moteur de bipèdes (Gorce, 1998; El Hafi, 2000). Les laboratoires de biomécanique se munissent d'instruments optoélectroniques pour étudier, par exemple, les mécanismes du mouvement de marche afin de concevoir des systèmes d'aide au diagnostic pour les cliniciens, la verticalisation de sujet pour concevoir des sièges électriques verticalisateurs (Gorce, 2003) ou encore l'ergonomie des sacs à dos (Gillet, 2004) et des postes de conduite automobile pour les industriels (Wang, 1999).

#### 4.5. Le choix du système d'analyse cinématique

Les systèmes optoélectroniques utilisant les marqueurs passifs sont les plus populaires dans l'analyse de la marche. La majeure partie des études sur la locomotion fait référence à ces systèmes comme étant les plus aptes à mesurer le mouvement de l'appareil locomoteur. Actuellement, il est possible d'enregistrer la trajectoire des marqueurs avec une fréquence comprise entre 60 et 1000 Hz. Le nombre de caméras peut être modulable, de quatre à seize voire davantage selon le système. Cette flexibilité permet l'acquisition de gestes complexes comme la marche pour éviter le plus possible la perte de vue des marqueurs par les caméras. De plus, les systèmes permettent d'enregistrer les trajectoires d'une grande quantité de marqueurs. Cette fonctionnalité favorise l'étude des gestes corporels avec un grand nombre de segments anatomiques pris en considération simultanément. Par conséquent, les modélisations du corps humain deviennent de plus en plus élaborées. Ainsi, l'utilisation de système VICON semble appropriée pour l'acquisition de la cinématique du genou lors de la marche mais il n'est pas utilisable en clinique.

##### 4.5.1. *Les systèmes de mesure dynamique*

L'analyse dynamique en biomécanique est l'étude des forces et moments internes qui engendrent le mouvement. Dans la pratique, il n'existe pas de méthode non invasive permettant une lecture directe des efforts transmis dans les articulations. Ainsi, le calcul des forces et moments internes passe par une modélisation du corps utilisant les lois de



la mécanique. En pratique, seuls les forces et moments externes peuvent être mesurés. Elles correspondent à la réaction du corps sur l'environnement. Lors de la marche, les pieds sont en interaction avec le sol. Dans ce cadre, les systèmes de mesure dynamique consistent à enregistrer les forces plus le centre de pression et moments externes de réaction du sol sur le pied. Les principaux outils utilisés pour mesurer ces interactions sont les capteurs de pression, les plates-formes de forces et les tapis de marche roulants.

#### *4.5.2. Les systèmes à capteurs de pression*

Placés sur une piste de marche, les tapis à capteurs de pression permettent l'enregistrement de la pression plantaire exercée par le pied sur le sol (exprimées en kilo pascal ( $1 \text{ N/cm}^2$ )). Ce dispositif utilise des cellules disposées sur la surface d'un tapis (avec environ 4 cellules au  $\text{cm}^2$ ). La technologie utilisée exploite des jauges de contrainte qui mesurent une force de pression proportionnelle à une différence de résistivité. Les capteurs peuvent être de types capacitifs ou à base de polymères conducteurs. L'ensemble des informations transmis par les capteurs permet de cartographier la répartition et l'intensité des pressions sous le pied pendant la phase d'appui. Il est alors possible de calculer le barycentre des pressions de la zone de contact (centre des pressions). Ainsi, l'évolution des positions du centre des pressions constitue la trajectoire verticale de réaction du sol sur le pied.

#### *4.5.3. Les plateformes de forces*

Les plates-formes de forces sont les outils dynamométriques les plus employés pour étudier la locomotion humaine. On obtient directement les forces et moments qui s'appliquent à l'interface entre le pied et la plate-forme dans les trois directions : verticale, longitudinale (antéro-postérieure) et latérale (médio-latérale). Ceux-ci sont mesurés au centre de la plateforme à l'aide d'un système de capteurs de force. Ces capteurs peuvent être des jauges de contrainte (AMTI, Logalex) ou des éléments piézoélectriques (Kistler). Les signaux issus des capteurs dérivent dans le temps et doivent par conséquent être remis à zéro à chaque essai.

#### *4.5.4. Les tapis de marche instrumentés*

Les tapis de marche instrumentés sont équipés de plates-formes de forces. Ces systèmes rendent possible l'enregistrement successif d'un grand nombre de mouvements de marche dans un espace restreint. Cependant, il faut noter deux considérations majeures. La première est liée au fait que le sujet marche sans avancer et que ses informations sensorielles s'en trouvent perturbées. La deuxième est une vitesse de marche imposée et donc implique des efforts supplémentaires de la part du sujet pour garder son équilibre. Ainsi, le tapis roulant modifie la vitesse naturelle de marche et peut modifier le pattern. En conséquence, le sujet s'accommode et stabilise sa foulée au régime de marche imposé par le tapis. Il est donc nécessaire que le sujet ait un temps d'adaptation. Plusieurs études ont été réalisées pour caractériser les modifications induites sur les

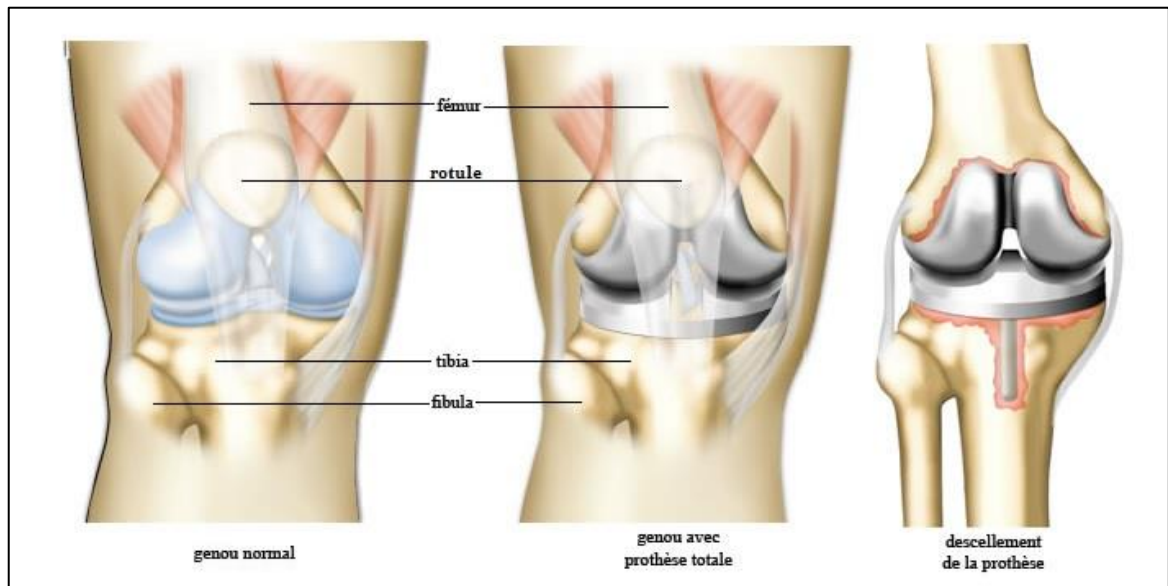
paramètres de marche. Toutefois, aucune conclusion n'a été apportée sur la validité des mesures : cinématique et dynamique réalisées avec un tapis roulant (Pélissier, 1994).

#### *4.5.5. Synthèse*

Les données cinématiques (trajectoires cartésiennes des marqueurs) proviennent d'un système développé pour une utilisation en milieu clinique. De plus c'est un système qui vient pallier au problème de précision de mesure des systèmes conventionnels qui utilisent des marqueurs collés directement sur la peau et ne permettent pas une mesure précise des rotations dans le plan transversal et frontal. En effet, l'acquisition de cycles répétitifs de marche en continu permet de limiter la variance intra-individuelle et améliorer la régularité des résultats.

### **5. L'arthroplastie du genou**

L'arthroplastie consiste à remplacer, en partie ou totalement, l'articulation malade par une prothèse. Cette chirurgie du genou permet de remplacer les surfaces endommagées de l'articulation afin de permettre une bonne mobilité ultérieure ([Figure 21](#)). Au Canada, il est réalisé annuellement en moyenne près de 55000 arthroplasties du genou avec en moyenne 10000 dans la province de Québec ([Tableau 1](#)).



**Figure 19 : Anatomie du genou vs arthroplastie du genou**

On distingue plusieurs types d'arthroplasties du genou :

### **5.1. L'arthroplastie partielle du genou (APG)**

Cette chirurgie est utilisée lorsqu'on doit changer une partie de l'articulation : on parle de l'arthroplastie partielle du genou (APG) ou prothèse unicompartmentale (PUC) de genou c'est-à-dire on remplace le côté latéral de l'articulation fémoro-tibiale (à la fois le fémur et le tibia)([Figure 22](#)).



**Figure 20 : Arthroplastie partielle du genou (APG)**

## 5.2. L'arthroplastie totale du genou (ATG)

L'arthroplastie totale du genou consiste à remplacer complètement toute l'articulation.

Ce choix est fait en fonction de l'état osseux et ligamentaire ([Figure 23](#)).



**Figure 21 : Arthroplastie totale du genou (ATG)**

Ce type de chirurgie améliore de façon considérable la mobilité des patients dont la maladie est très avancée que les soins médicaux ne donnent plus des satisfactions ou bien de maladies inflammatoires.

### **5.3. Les fondements de l'intervention de l'arthroplastie du genou**

Les critères sur lesquels peuvent fonder la décision de se faire opérer (Viel, 2000) du patient sont:

- Les douleurs handicapantes malgré le traitement médical bien conduit;
- La mobilité du genou;
- Les difficultés grandissantes pour monter et descendre des marches ;
- État avancé du cartilage endommagé et ses conséquences sur les ligaments;
- Résultat complémentaire des examens de radiographies, Scanner et/ou IRM.

### **Conclusion**

Dans ce chapitre nous avons décrit une mise en contexte de la biomécanique du genou, du cycle de marche, de la cinématique du genou et les systèmes de mesure du mouvement pour mieux comprendre la problématique liée à nos travaux de recherches. Nous avons démontré que les diagnostics des maladies liées à l'articulation du genou demeurent empiriques et reposent sur le seul jugement de l'expert médical.

Il convient donc d'uniformiser les connaissances pour pouvoir les exploiter dans des champs plus larges. Pour parvenir à ce niveau cognition, il est nécessaire de mettre en place des protocoles de mesure et des outils d'analyse pour quantifier les paramètres biomécaniques pertinents permettant l'évaluation objective des défauts du mouvement, puis de proposer une solution de correction adaptée. Les données biomécaniques obtenues peuvent être cinétiques tridimensionnelles (3D), cinématiques et accélérométriques à proximité du centre de gravité du genou, à savoir le déplacement angulaire du tibia versus du fémur en fonction du cycle de marche. Ainsi, nous proposons d'étudier les paramètres biomécaniques identifiés dans ce chapitre et impliqués dans la pathologie : gonarthrose. En effet, il est nécessaire de comparer ces paramètres sur des sujets suivant deux modalités : genou arthrosique non chirurgical et genou arthrosique chirurgical. Les variations entre les deux modalités nous permettront d'identifier et de quantifier le dysfonctionnement et le fonctionnement « optimal » de l'appareil locomoteur et enfin de diagnostiquer la pathologie: l'arthroplastie. Nous serons ensuite en mesure d'étudier les interrelations entre les paramètres biomécaniques du membre inférieur.

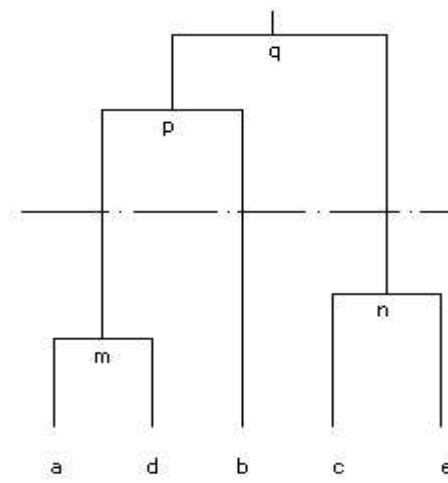
## Chapitre 3

### La classification des formes

La classification de formes consiste à grouper des éléments d'un ensemble d'individus ou d'objets selon des classes (aussi appelées groupes ou catégories). Les éléments d'une classe sont des individus qui possèdent des caractéristiques communes.

L'objectif de la classification est de produire une structure, une arborescence qui met en évidence les liens hiérarchiques entre les individus ou entre des groupes d'individus. Par exemple, sur l'arbre hiérarchique ci-dessous (figure 22), les individus a et d sont très proches et sont aussi assez proches de l'individu b. Ces trois individus forment une classe. D'autre part, les individus c et e se ressemblent et forment aussi une classe. Au-delà de ces sous classes, les cinq individus se ressemblent encore plus et constituent une grande classe. Cette représentation sous forme d'arbre permet également de détecter un nombre de classes naturel dans une population. Par exemple dans l'arbre proposé, on peut décomposer les individus en trois classes, la classe a-d, la classe b et la classe c-e.





**Figure 22 : Arbre hiérarchique constitué de cinq individus a-b-c-d-e et formant trois classes.**

Pour réaliser une classification, il faut définir une mesure de ressemblance entre deux individus. Ce qui nous amène à se poser des questions suivantes : Quand est-ce que deux individus sont très proches ? Quand est-ce qu'on va les mettre dans une même classe ?

De nos jours, la problématique n'est plus de reproduire les données du mouvement biomécanique de l'être humain, mais surtout de savoir comment les utiliser afin qu'elles soient cliniquement interprétables (Whittle, 2003). L'introduction des données biomécanique dans le domaine médicale pour l'évaluation et l'aide à la décision clinique non invasive comporte quelques contraintes : (1) les données sont de grande

dimension et difficiles à interpréter par les cliniciens (Deluzio et Astephen, 2007; Jones et al., 2006; Jones, Holt et Beynon, 2008) et (2) la classification automatique des données consiste à traiter les données pour trouver les informations cachées qui pourront amener à affecter chaque sujet à sa propre classe.

On distingue deux types de classifications : La classification supervisée et la classification non supervisée.

## **1. La classification supervisée**

Dans la classification supervisée, les méthodes les plus connues sont les suivantes:

### **1.1. Les k-plus proches voisins**

Encore appelée en anglais la méthode K-Nearest Neighbors (K-NN), consiste à déterminer la liste des k plus proches voisins parmi les individus déjà classés pour chaque nouvel individu à classer ; k étant le nombre de voisins à prendre en compte. L'individu est affecté à la classe qui contient le plus d'individus parmi ces k plus proches voisins. La distance la plus utilisée est la distance Euclidienne (Viel, 2000).

Soient deux individus  $(i, k)$  caractérisés par le vecteur de caractéristique  $x$  de dimension  $p$ . La distance euclidienne entre ces individus est obtenue par la formule:

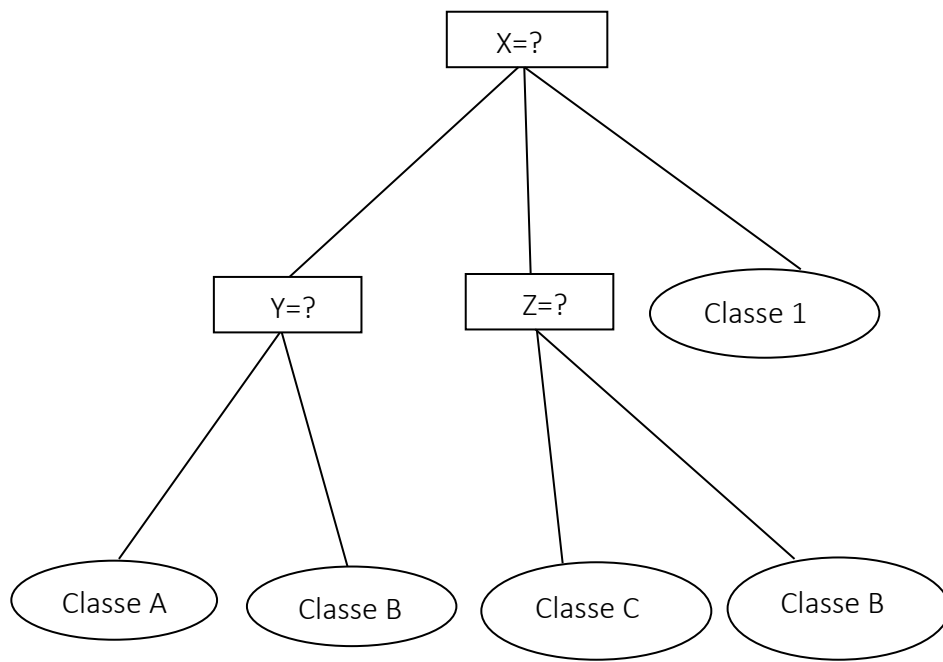
$$d^2(i, k) = \sum_{j=1}^p (x_{ij} - x_{kj})^2$$

## 1.2. Les arbres de décision

Un arbre de décisions (decision tree) est une structure très utilisée en classification de formes. Son fonctionnement repose sur des heuristiques construites selon des techniques d'apprentissage supervisé (Mezghani, 2015). C'est un classificateur interprétable représenté sous forme d'arbre tel que :

- Les nœuds de l'arbre testent les attributs ;
- Chaque nœud réalise un test portant sur la valeur d'un attribut dont le résultat indique la branche à suivre dans l'arbre ;
- Il y a une branche pour chaque valeur possible de l'attribut testé ;
- Les feuilles spécifient les catégories (deux ou plus).

Contrairement au K-NN, les arbres de décision fonctionnent facilement sur les données qualitatives tant que le nombre de caractéristiques n'est pas trop grand.



**Figure 23 : Exemple d'arbre de décision avec les classes qui sont les feuilles**

Construire un tel arbre consiste à définir un nœud, chaque nœud permettant de faire une partition des objets en 2 groupes sur la base d'une des variables explicatives. Il convient donc :

- De définir un critère permettant de sélectionner le meilleur nœud possible à une étape donnée;

- De définir quand s'arrête le découpage, en définissant un nœud terminal (feuille);
- D'attribuer au nœud terminal la classe ou la valeur la plus probable;
- D'élaguer l'arbre quand le nombre de nœuds devient trop important en sélectionnant un sous arbre optimal à partir de l'arbre maximal;
- Valider l'arbre à partir d'une validation croisée ou d'autres techniques.

Pour construire un arbre de décision, on se sert de plusieurs critères tels que le gain d'information et l'entropie.

Le gain d'information est donné par :

$$Gain(X, A) = E(A) - \sum_{j=1}^m \frac{|A_j|}{|A|} E(A_j)$$

Avec :

- $A$  l'ensemble de données
- $m$  le nombre de valeurs que peut prendre l'attribut  $X$  considéré
- $A_j$  une partition donnée de  $A$

L'entropie de la classification doit être calculée au préalable. Elle est définie par :

$$E(P) = - \sum_{i=1}^n p_i \log p_i \text{ où } p_i = \frac{|C_i|}{T}$$

Avec :

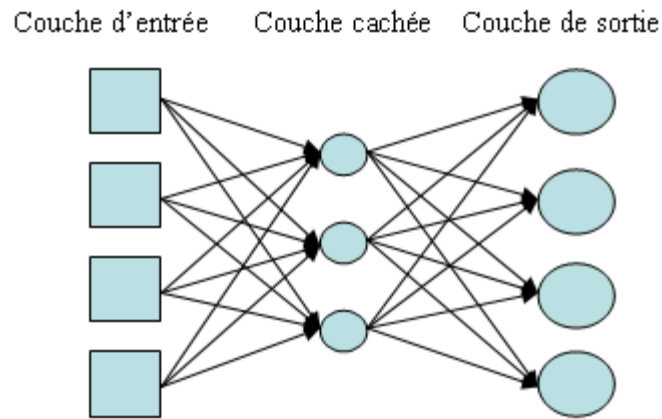
- $p_i$  la probabilité de la classe  $i$
- $|C_i|$  désigne le nombre d'éléments en valeur absolue de la classe  $i$
- $T$  l'ensemble de données

L'entropie permet de donner la façon standard de quantifier l'information moyenne contenue par une observation  $x$ . Plus  $p(x)$  est proche d'une loi uniforme, plus l'entropie est élevée.

Si  $p(x)=1$  pour une seule valeur de  $x$ , l'entropie vaut 0.

### 1.3. Les réseaux neurones

Un réseau de neurones est une méthode de calcul à base de fonctions mathématiques dont la conception est inspirée du fonctionnement des neurones biologiques. En effet les réseaux de neurones sont une méthode de l'intelligence artificielle optimisée par des méthodes d'apprentissage de type probabiliste.



**Figure 24 : Réseau de neurones**

#### 1.4. Naïve bayes

Un réseau bayésien est un graphe orienté et acyclique qui contient des nœuds connectés entre eux par des arcs représentant des dépendances probabilistes entre les variables et des distributions de probabilités conditionnelles pour chaque variable. La classification naïve bayésienne est basée sur le théorème de Bayes avec une forte indépendance des hypothèses.

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

### 1.5. Les machines à support de vecteurs

Introduites récemment par Vapnik (Vapnik, 2000), les machines à support de vecteurs sont un algorithme qui consiste à résoudre les problèmes de discrimination qui regroupent deux classes. On essaie de déterminer la classe à laquelle appartient un individu entre deux choix possibles. Autrement dire, les classes sont séparées par un hyper-plan optimal qui sépare correctement toutes les données et qui maximise la marge, la distance du point le plus proche à l'hyper-plan (représentée par  $d$  dans la [figure 27](#)).

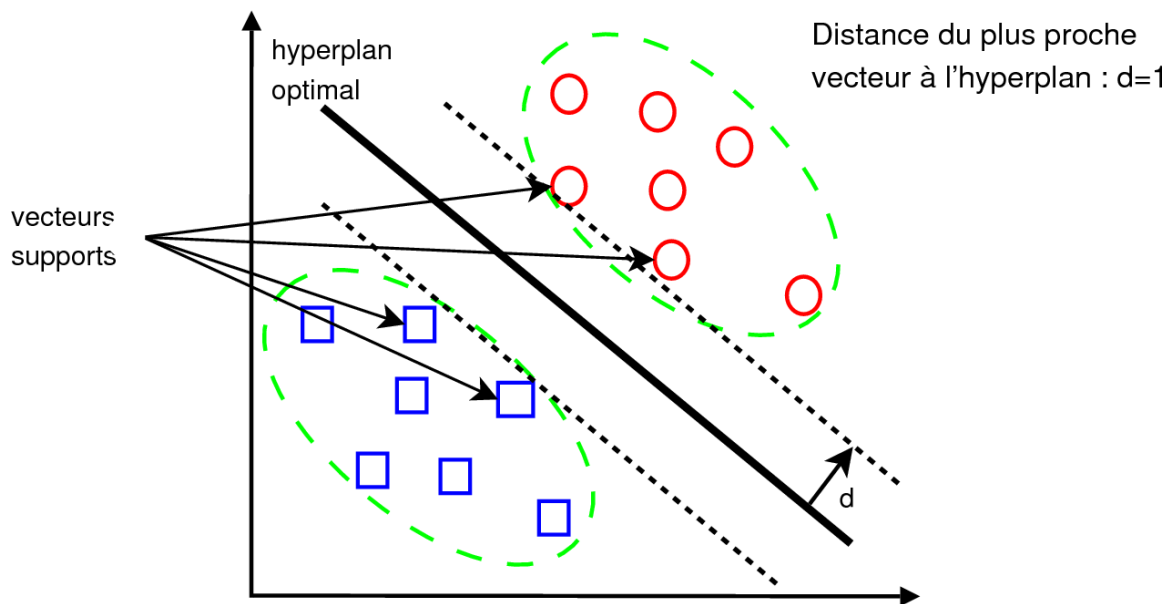


Figure 25 : Hyper-plan optimal et marge maximale



## 2. La classification non-supervisée

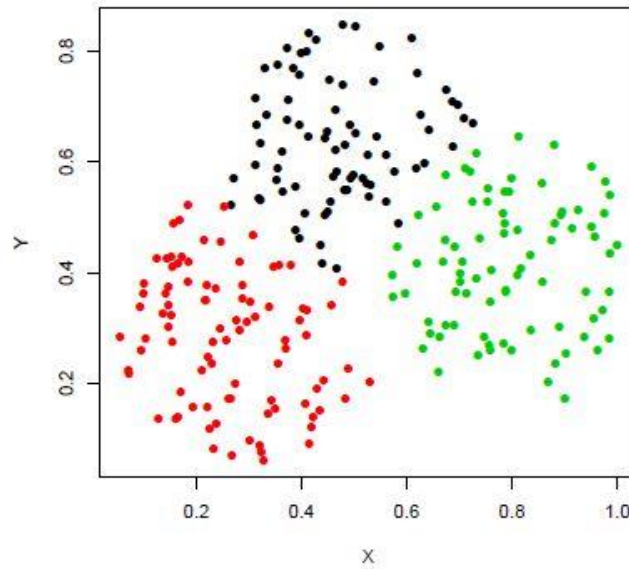
La classification non-supervisée est un "clustering" ou un "regroupement" qui consiste à identifier les populations d'un ensemble de données. Par exemple, supposons qu'on dispose d'un ensemble d'objets noté  $X = \{x_1, x_2, \dots, x_N\}$  qui caractérise un ensemble de descripteurs  $D$ , l'objectif du clustering est d'obtenir les ensembles auxquels appartiennent chaque objet  $x$  noté  $C = \{C_1, C_2, \dots, C_n\}$ . Ce qui revient à déterminer une fonction notée  $Y_s$  qui associe à chaque élément de  $X$  un ou plusieurs éléments de  $C$ .

Les deux méthodes essentielles les plus utilisées dans les classifications non-supervisées sont : l'approche des centres mobiles (k-means) et l'approche de la classification hiérarchique (Vapnik, 2000).

### 2.1. L'approche du K-moyenne

L'algorithme des k-moyennes (en anglais  $k$ -mean), est le plus connu des algorithmes non hiérarchiques. Il utilise la distance euclidienne pour des variables quantitatives. Il consiste à minimiser la somme des carrés des écarts aux centroides de chaque classe. Le nombre  $k$  de classes est supposé ici connu d'avance. Par exemple, pour un tableau de données  $(x_{ij})_{i=1, \dots, n, j=1, \dots, d}$ , l'algorithme des  $k$ -means cherche à trouver la partition  $C_1, \dots, C_k$  qui minimise le critère d'inertie intra-classes.

$$\sum_{j=1,\dots,k} \sum_{i \in C_j} \|x_i - p_j\|^2 \text{ où } p_j = \frac{1}{|C_j|} \sum_{i \in C_j} x_i$$



**Figure 26 : Partitionnement de données avec l'algorithme K-moyennes.**

Itérativement l'algorithme procède à l'alternance de l'affectation des observations à la classe dont le centre de classe est le plus proche avec le re-calcule des centres de classe. La convergence de l'algorithme est généralement atteinte en quelques itérations. Il existe deux méthodes connues sur le principe de k-means : la méthode de centres mobiles et la méthode des nuées dynamiques.

### *2.1.1. Méthode des centres mobiles*

Cette méthode consiste à construire une partition en  $k$  classes en sélectionnant  $k$  individus commence, des classes tirées au hasard de l'ensemble d'individus. Après cette sélection, on affecte chaque individu au centre le plus proche en créant  $k$  classes, les centres des classes seront remplacées par les centres de gravité et nouvelles classes seront créées par le même principe. Généralement la partition obtenue est localement optimale car elle dépend du choix initial des centres. Pour cela les résultats entre deux exécutions de l'algorithme sont significativement variés.

### *2.1.2. Méthode des nuées dynamiques*

Dans ce cas-ci, un seul point ne représente pas une classe mais plutôt un noyau de points qui constitue les éléments représentatifs de la classe. Cela permet de corriger l'influence d'éventuelles valeurs extrêmes sur le calcul du barycentre.

## **2.2. La classification hiérarchique**

C'est un ensemble de méthodes qui permettent de classer des individus, des modalités, des variables en divers groupes que l'on souhaitera homogènes pour les variables que l'on soumet à l'analyse. Chacune de ces méthodes conduit à la réalisation d'un arbre hiérarchique à partir duquel l'analyste décidera du nombre de groupes à constituer. Le

processus de classification est itératif. Il est construit d'une part sur les distances entre individus et d'autre part entre groupes d'individus.

On distingue deux types de classifications hiérarchiques : La classification hiérarchique ascendante et la classification hiérarchique descendante.

Étant donné que la deuxième est moins utilisée, nous présenterons donc ici la première approche.

### *2.2.1. La classification hiérarchique ascendante*

La classification hiérarchique ascendante permet de construire une hiérarchie entière des objets sous la forme d'un "arbre" dans un ordre ascendant. Au départ, on considère chaque individu comme une classe et on essaye de fusionner deux ou plusieurs classes appropriées (selon une similarité) pour former une nouvelle classe. Le processus est itéré jusqu'à ce que tous les individus se trouvent dans une même classe. On génère ainsi un arbre qu'on peut intercepter à différents niveaux pour obtenir un nombre des classes.

Il existe de différentes mesures de la distance interclasses qui peuvent être utilisées : la distance euclidienne, la distance inférieure ou la distance supérieure.

Pour qu'un groupe d'individus soit homogène, il est nécessaire de mesurer une ressemblance entre deux observations.

### 3. Le bagging et le boosting

Les méthodes d'ensemble constituent une famille ou ensemble d'algorithmes qui génèrent une collection de classificateurs pour par la suite les combiner en agrégeant leurs prédictions (Vapnik, 2000). L'efficacité de la combinaison des classificateurs repose principalement sur leur capacité à tirer les complémentarités des classificateurs individuels dans le but d'améliorer autant que possible les performances (capacité de prédiction) en généralisation de l'ensemble. D'une manière générale, une méthode d'ensemble est caractérisé par :

- un espace d'hypothèses  $H$
- un algorithme d'apprentissage  $A$  produisant les hypothèses  $h_1, \dots, h_t \in H$
- l'agrégation des  $h_n$  par vote de majorité pour la classification

Le vote de majorité consiste à chercher la classe la plus majoritaire parmi les classes prédites par les classificateurs de la collection et de retourner cette classe comme la classe du vote de majorité.

Plusieurs auteurs ont déjà démontré l'utilité des méthodes d'ensemble dans l'amélioration des performances des classificateurs qui au départ donnaient de mauvaises performances. Dans les sections suivantes, nous allons introduire quelques méthodes d'ensemble couramment utilisées en apprentissage automatique : le Bagging

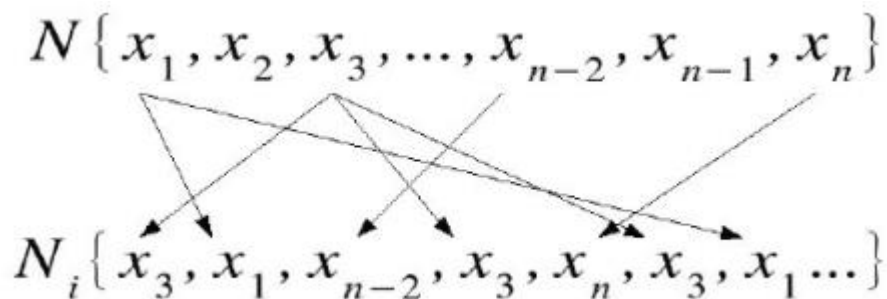
et le Boosting. Le tableau 3 contient un récapitulatif de la comparaison de ces deux méthodes.

### 3.1. Le bagging

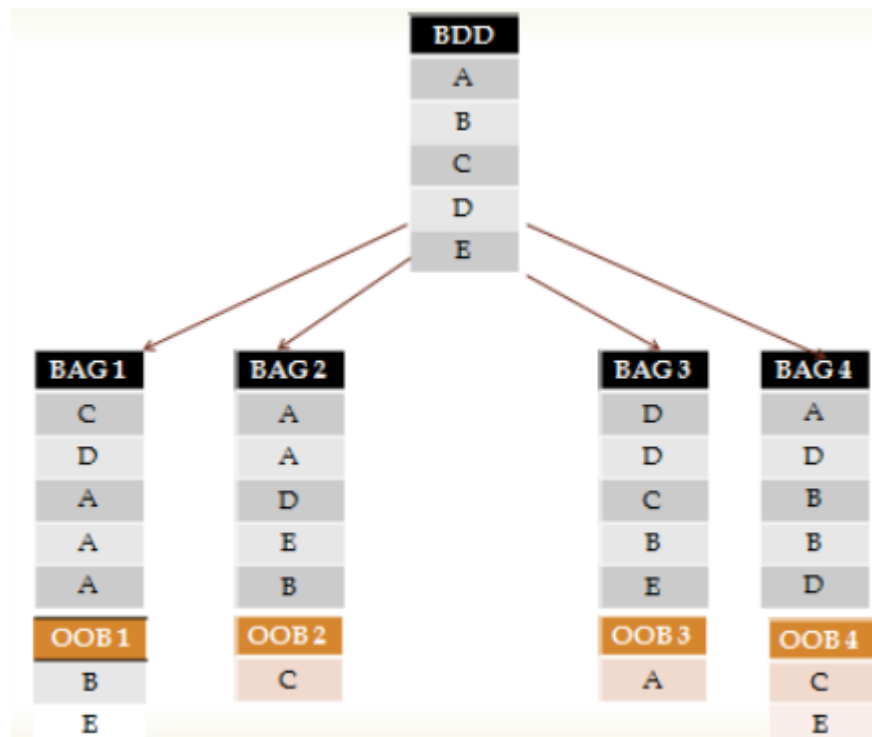
Le « bagging » est un acronyme qui provient de « **bootstrap aggregating** ». C'est un grand algorithme qui combine la technique de «bootstrap» et celle de «aggregating». C'est un principe d'apprentissage d'ensemble de classifieurs qui peut être utilisé avec tout type de classifieur élémentaire, mais dont l'efficacité est démontrée par Breiman principalement dans le cadre de combinaison d'arbres de décision. L'idée de base est d'entraîner un algorithme d'apprentissage élémentaire sur plusieurs bases d'apprentissage obtenues par tirage avec remise.

### 3.2. Bootstrap

Si on considère un ensemble  $N$ , le bootstrap de cet ensemble  $N$ , est l'ensemble des données obtenues à la suite d'un tirage aléatoire  $n$  fois des éléments de  $N$  avec remise (Breiman, 2000).



**Figure 27: Illustration d'un tirage aléatoire avec remise pour la formation d'un échantillon**

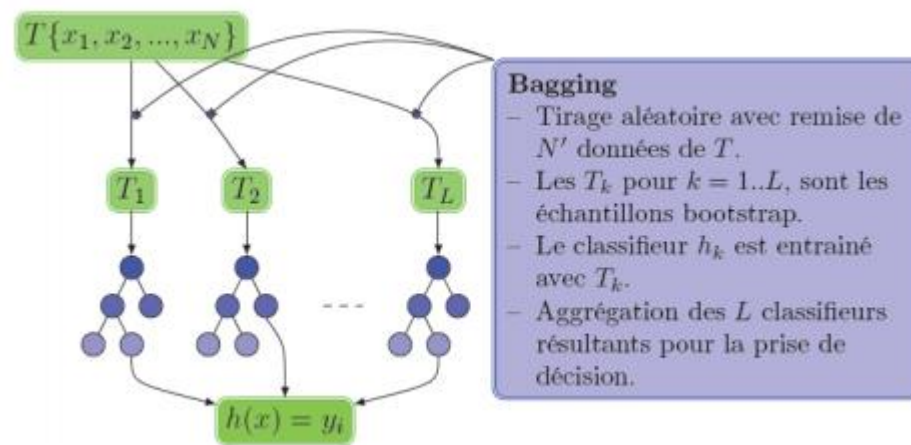


**Figure 28: Des échantillons bootstrap d'une base de données**

### 3.3. Agrégation

Elle consiste à faire l'échantillonnage avec remplacement de l'ensemble d'apprentissage; Ainsi, une observation peut apparaître plusieurs fois dans un même

échantillon (bag) et chaque bag contient environ 2/3 des observations de l'ensemble d'apprentissage et environ 1/3 des observations resteront en dehors du bag (out-of-bag ou OOB). L'algorithme construit des centaines d'arbres décisionnels et les combine en un seul modèle. Dans le cas d'une classification, on prendra comme résultat le vote majoritaire et si c'est une régression, on considèrera la moyenne.



**Figure 29: Fonctionnement de la technique de Bagging**



#### 4. Le boosting

Le Boosting est une technique qui consiste à optimiser les performances de nombreux algorithmes regroupés qui utilisent des ensembles de classifieurs binaires afin de fournir des décisions très précises. (Freund et Schapire, 1996). Tout comme le Bagging, le Boosting combine aussi les apprentissages réalisés pour en sortir une règle plus efficace de la classification mais il est plus performant que ce dernier.

Bagging	Boosting
Aléatoire	Adaptatif et généralement déterministe
Utilise des échantillons Bootstrap	Utilise échantillon initial au complet
Les modèles ont le même poids	Les modèles pondérés selon leur qualité d'ajustement

**Tableau 3: Bagging vs Boosting**

## Chapitre 4

### Classification des données biomécaniques : Méthodologie et résultats expérimentaux

Notre étude porte, rappelons-le, sur le développement d'un système de classification pour classer des patients ayant une arthrose au genou selon deux classes : une classe de sujets chirurgicaux pour une arthroplastie et une classe de sujets non chirurgicaux. La classification utilise les données cinématiques du patient enregistrées dans les plans suivants : frontal, sagittal et transversal. Ces données correspondent respectivement au mouvement d'abduction/adduction, de flexion/extension et de rotation interne et externe durant une tâche de marche normale, une vitesse de marche confortable sur tapis roulant. Pour ce faire, nous avons développé un système de classification basé sur les forêts aléatoires (section 2). Le système de classification a été testé sur une base de données de 89 sujets gonarthrosiques repartitionnés en deux groupes : 48 sujets chirurgicaux (Ch) et 41 sujets non chirurgicaux (NCh) (section 1). Les forêts aléatoires sont testées plan par plan de rotation et seront ensuite comparées en regardant les matrices de confusion. Les performances du système de classification ont été évaluées en termes de taux de bonne classification, de sensibilité et spécificité (section 4). Dans ce qui suit nous avons détaillé chacune de ces modules du système de classification.

## 1. Base de données

Notre base de données contient les données cinématiques de 89 sujets gonarthrosiques repartitionnés en deux groupes : 48 sujets chirurgicaux (Ch) et 41 sujets non chirurgicaux (NCh). Les sujets non chirurgicaux avaient un diagnostic d'arthrose et présentaient les signes cliniques associés tandis que les sujets chirurgicaux doivent être exclus de toutes autres conditions qui pourraient influencer l'évaluation de la marche ou la sécurité de l'évaluation. Les différentes variables décrivant ici les sujets de notre base de données étant les variables d'intérêt, c'est-à-dire les variables quantitatives, il est donc possible d'utiliser toutes les méthodes de classification qui existent, mais ici nous n'en présenterons que la forêt aléatoire. Les caractéristiques démographiques descriptives des participants sont présentées dans le tableau 4.

	Sujets Chirurgicaux (Ch)	Sujets Non Chirurgicaux (NCh)
Effectifs	48	41
Age (an)	64.62±8.88	64.12±9.24
Taille (cm)	166.7±9.76	169.4±8.40
Poids (Kg)	93.3±25.41	89.7±19.91

Pourcentage homme/femme	62.5/37.50	56.10/43.9
----------------------------	------------	------------

**Tableau 4 : Caractéristiques démographiques et descriptives des participants**

## 2. Les forêts aléatoires

Les forêts aléatoires sont composées (comme le terme "forêt" l'indique) d'un ensemble d'arbres décisionnels binaires dans lequel a été introduit de l'aléatoire. Ces arbres se distinguent les uns des autres par le sous-échantillon de données sur lequel ils sont entraînés. Ces sous-échantillons sont tirés au hasard (d'où le terme "aléatoire") dans un jeu de données.

La technique des forêts aléatoires modifie la méthode du Bagging appliquée ici aux arbres en ajoutant un critère de dé-corrélation entre ces arbres. L'idée de cette méthode est de réduire la corrélation sans augmenter trop la variance. Le principe consiste à choisir de façon aléatoire un sous-ensemble de variables qui sera considéré à chaque niveau de choix du meilleur nœud de l'arbre.

## 2.1. Principe et algorithme des forêts aléatoires

Considérons un ensemble d'entraînement  $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$ ,  $a$  le nombre d'attributs des exemples de  $X$ . Considérons également  $S_t$  un bootstrap (voir Section 2.1) contenant  $m$  instances obtenus par rééchantillonnage avec remplacement de  $S$ . Soit  $\{h_1, \dots, h_T\}$  un ensemble de  $T$  arbres de décision. Chaque arbre  $h_t$  est construit à partir de  $S_t$ . Pour chaque nœud de l'arbre, l'attribut de partitionnement est choisi en considérant un nombre  $f(f < a)$  d'attributs choisis aléatoirement (parmi les  $a$  attributs). Pour classifier une nouvelle instance  $x$ , le classificateur des forêts aléatoires effectue un vote de majorité uniformément pondéré des classificateurs de cet ensemble pour l'instance  $x$ . L'algorithme illustre ce principe.

**Algorithme :** RandomForest( $S, T$ )

Entrée :  $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$ , l'ensemble d'entraînement.

Entrée :  $T$ , le nombre d'arbres de décision de la forêt aléatoire.

**Pour**  $t = 1, \dots, T$  faire

1. Générer un échantillon Bootstrap  $S_t$  de taille  $m$  à partir de  $S$

2. Créer un arbre de décision  $h_t$  à partir de  $S_t$  en répétant de la manière récursive pour chaque nœud de l'arbre les étapes suivantes :

- a) Sélectionner aléatoirement  $f$  attributs parmi les  $a$  attributs
- b) Choisir l'attribut de partitionnement parmi les  $f$
- c) Partitionner le nœud en deux nœuds fils

Fin pour

Sortie :  $H$ , le classificateur de la forêt aléatoire

Le principal avantage de cet algorithme est qu'il permet d'éviter le danger que représente le sur-apprentissage pour toute méthode de prédiction basée sur l'induction. BREIMAN, 2001 démontre que lorsque le nombre d'arbres impliqués dans la forêt de prédiction augmente, le taux d'erreur en généralisation converge vers une valeur limite, dont une borne supérieure peut être estimée sur une base des caractéristiques intrinsèques de la forêt. Il s'agit de la propriété de convergence des forêts aléatoires. Cela explique pourquoi les forêts aléatoires ne font pas de sur-apprentissage lorsque le nombre d'arbres de la forêt augmente mais plutôt convergent vers une valeur limite de l'erreur OOB.

## 2.2. L'erreur en dehors du bootstrap ou l'Out-Of-Bag (OOB)

C'est l'ensemble des exemples qui ne sont pas sélectionnés dans les échantillons bootstrap. C'est un paramètre qui permet l'évaluation interne du classifieur et l'estimation de l'importance des variables pour la sélection de variables.

## 3. Les critères d'évaluation

Les performances de classification des données ont été évaluées par le calcul des vrais positifs (VP), des vrais négatifs (VN), des faux positifs (FP) et des faux négatifs (FN), le pourcentage de sensibilité (Se), la spécificité (Sp) et le taux de classification (Tc), leurs définitions respectives sont les suivantes :

- VP : Vrai Positif : nombre de positifs classés positifs.
- VN : Vrai Négatif : nombre de négatifs classés négatifs.
- FP : Faux Positif : nombre de négatifs classés positifs.
- FN : Faux négatif : nombre de positifs classés négatifs.

La sensibilité : c'est la capacité d'obtenir un résultat positif alors que la maladie est présente, elle est calculée par :

$$S_e = \frac{VP}{VP+FN}$$

La spécificité : c'est la capacité d'obtenir un résultat négatif alors que la maladie est absente, elle est calculée par :

$$S_p = \frac{VN}{VN+FP}$$

Le taux de classification : c'est le pourcentage des exemples correctement classés, il est calculé par :

$$T_c = \frac{VP+VN}{VP+VN+FP+FN}$$

Les performances du système de classification sont aussi évaluées en utilisant la matrice de confusion qui est un outil performant et bien adapté aux problèmes de classification. Les lignes de la matrice de confusion représentent les prédictions alors que les colonnes représentent les classes réelles, dans notre cas, nous avons une matrice binaire de 2x2. Nous considérons un individu chirurgical (1) comme négatif et un sujet non chirurgical comme positif (0).

La matrice de confusion est obtenue grâce à l'ensemble de donnée de test. Celle-ci est passée dans le modèle prédictif et la variable à prédire est comparée à la variable obtenue par le modèle.



Par exemple, si le prédicteur donne 1 à la sortie pour un individu prédit alors que dans la donnée de test la classe de l'individu en question est 0, il s'agit d'un faux négatif. Grâce à la matrice de confusion, nous pouvons calculer les différents taux.

#### **4. Résultats expérimentaux et discussion**

Nous avons utilisé l'algorithme des forêts aléatoires introduit par Leo Breiman. L'algorithme des forêts aléatoires utilise une distribution uniforme sur les arbres de la forêt dans le vote de majorité. Nous nous sommes particulièrement intéressés à Tree Bagging tels que définie par Breiman (Breiman L., 1996).

Les simulations ont été réalisées dans l'environnement de programmation statistique MATLAB R2016a version 9.0.0.341360 64-bit (win64) sous Windows 10 Professionnel. La procédure correspondant aux implémentations des méthodes de classification et forêt aléatoire dans l'environnement MATLAB est la fonction Treebagger.

Pour garantir une estimation non biaisée des résultats de classement obtenus par la méthode présentée, le jeu de données initial a préalablement été scindé en deux ensembles dans les proportions 70% et 30% respectivement pour l'ensemble d'apprentissage et pour l'ensemble de test par un tirage aléatoire sans remise des individus. Un jeu de données indépendant de 63 observations a été utilisé pour

construire l'estimateur (échantillon d'apprentissage), les 26 observations restantes sont donc réservées pour l'estimation de l'erreur de classement (échantillon test). L'ensemble d'apprentissage est constitué de 30 sujets chirurgicaux et de 33 sujets non chirurgicaux. Quant à L'ensemble de test, il est constitué de 18 sujets chirurgicaux et de 8 sujets non chirurgicaux (Tableau 4).

	Apprentissage	Test	Total
Sujets chirurgicaux	30	18	48
Sujets Non chirurgicaux	33	8	41
Total	63	26	89

**Tableau 4 : Répartition de la BD en jeux de données**

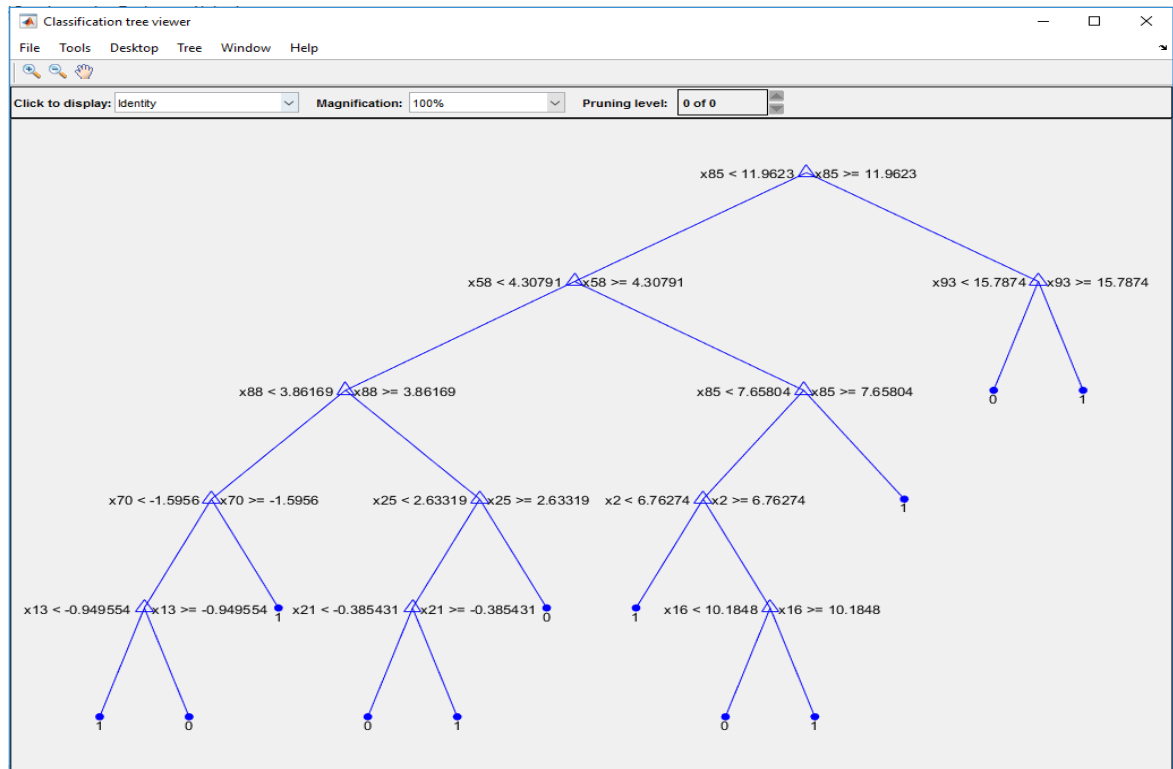
La méthode consiste à choisir de façon aléatoire un sous-ensemble de variables qui sera considéré à chaque niveau de choix du meilleur nœud de l'arbre. L'algorithme génère un jeu d'arbres doublement perturbés au moyen d'une randomisation opérée à la fois au niveau de l'échantillon d'apprentissage et des partitions internes.

Chaque arbre du jeu est ainsi généré au départ d'un sous-échantillon bootstrap du jeu d'apprentissage complet, de manière similaire aux techniques de bagging. Ensuite, l'arbre est construit en utilisant la méthodologie CART, à la différence près qu'à chaque nœud la sélection de la meilleure partition s'effectue non pas sur les 100 variables au complet mais sur un sous-ensemble sélectionné aléatoirement au sein de celles-ci. La taille de notre forêt (nTrees) est fixée préalablement à l'exécution de la fonction. Chaque arbre est ainsi développé jusqu'à sa taille maximale, sans élagage.

C'est cette méthode qui est alors appliquée sur le jeu d'apprentissage avec les paramètres par défaut de la fonction **TreeBagger(nTrees, features, classLabels, 'oobPred', 'On', 'Method', 'classification')** dans laquelle nous définissons le nombre d'arbres (nTrees=10) dans notre forêt.

Sur la base des estimateurs ainsi obtenus, une prédiction des deux classes de patients a été effectuée sur le jeu de test. Le programme MATLAB présentant la fonction TreeBagger, qui nous a permis d'obtenir les informations caractéristiques de notre analyse des données est en Annexe A.

Le graphe du premier arbre de notre forêt construit avec notre modèle est obtenu avec la fonction **view(B.Trees{1}, 'mode', 'graph')** et sa description est obtenue avec la fonction **view(B.Trees{1})** tandis que celui du dernier arbre est **view(B.Trees{nTrees}, 'mode', 'graph')** et sa description est **view(B.Trees{nTrees})**.



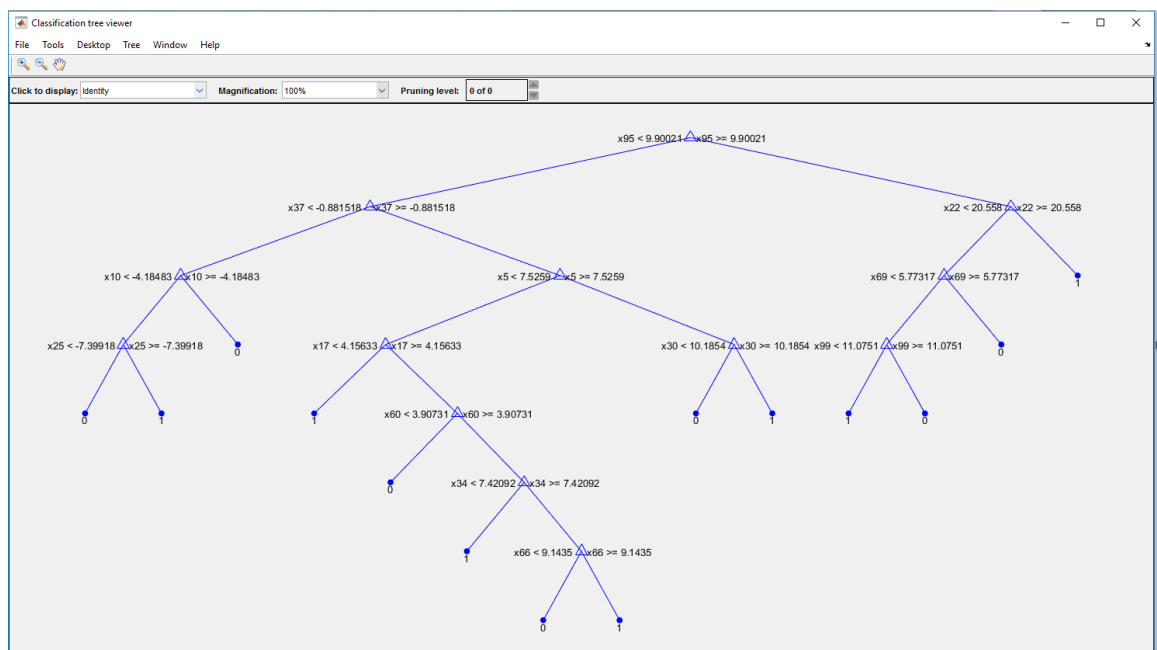
**Figure 30 : Premier arbre de décision des 10 arbres de la forêt des données d'abduction-adduction**

```

Decision tree for classification
1  if x85<11.9623 then node 2 elseif x85>=11.9623 then node 3 else 1
2  if x58<4.30791 then node 4 elseif x58>=4.30791 then node 5 else 1
3  if x93<15.7874 then node 6 elseif x93>=15.7874 then node 7 else 0
4  if x88<3.86169 then node 8 elseif x88>=3.86169 then node 9 else 0
5  if x85<7.65804 then node 10 elseif x85>=7.65804 then node 11 else 1
6  class = 0
7  class = 1
8  if x70<-1.5956 then node 12 elseif x70>=-1.5956 then node 13 else 1
9  if x25<2.63319 then node 14 elseif x25>=2.63319 then node 15 else 0
10 if x2<6.76274 then node 16 elseif x2>=6.76274 then node 17 else 1
11 class = 1
12 if x13<-0.949554 then node 18 elseif x13>=-0.949554 then node 19 else 0
13 class = 1
14 if x21<-0.385431 then node 20 elseif x21>=-0.385431 then node 21 else 1
15 class = 0
16 class = 1
17 if x16<10.1848 then node 22 elseif x16>=10.1848 then node 23 else 0
18 class = 1
19 class = 0
20 class = 0
21 class = 1
22 class = 0
23 class = 1

```

**Figure 31 : Description du premier arbre de décision des 10 arbres de la forêt des données d'abduction-adduction**



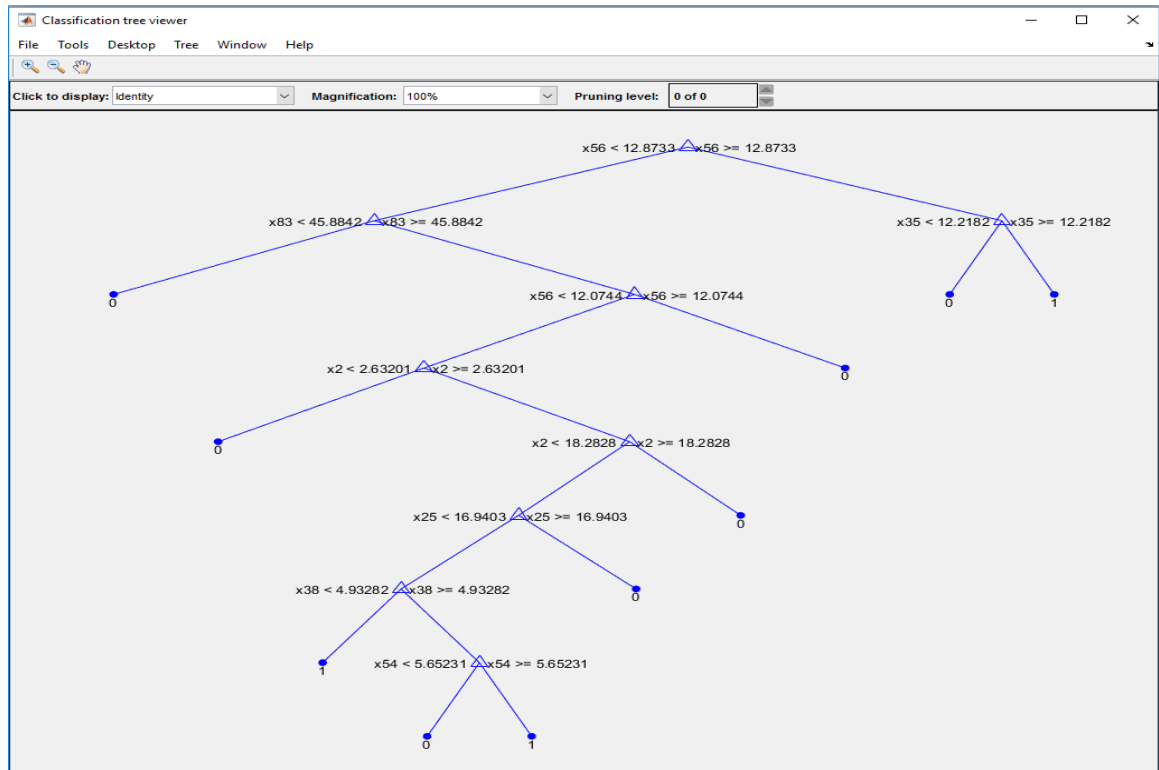
**Figure 32 : Dernier (10ème) arbre de décision des 10 arbres de la forêt d'abduction-adduction**

```

Decision tree for classification
1  if x95<9.90021 then node 2 elseif x95>=9.90021 then node 3 else 0
2  if x37<-0.881518 then node 4 elseif x37>=-0.881518 then node 5 else 0
3  if x22<20.558 then node 6 elseif x22>=20.558 then node 7 else 0
4  if x10<-4.18483 then node 8 elseif x10>=-4.18483 then node 9 else 0
5  if x5<7.5259 then node 10 elseif x5>=7.5259 then node 11 else 1
6  if x69<5.77317 then node 12 elseif x69>=5.77317 then node 13 else 0
7  class = 1
8  if x25<-7.39918 then node 14 elseif x25>=-7.39918 then node 15 else 0
9  class = 0
10 if x17<4.15633 then node 16 elseif x17>=4.15633 then node 17 else 1
11 if x30<10.1854 then node 18 elseif x30>=10.1854 then node 19 else 0
12 if x99<11.0751 then node 20 elseif x99>=11.0751 then node 21 else 0
13 class = 0
14 class = 0
15 class = 1
16 class = 1
17 if x60<3.90731 then node 22 elseif x60>=3.90731 then node 23 else 1
18 class = 0
19 class = 1
20 class = 1
21 class = 0
22 class = 0
23 if x34<7.42092 then node 24 elseif x34>=7.42092 then node 25 else 1
24 class = 1
25 if x66<9.1435 then node 26 elseif x66>=9.1435 then node 27 else 0
26 class = 0
27 class = 1

```

**Figure 33 : Description du dernier arbre de décision des 10 arbres de la forêt des données d'abduction-adduction**



**Figure 34 : Premier arbre de décision des 10 arbres de la forêt des données de Flexion-extension**

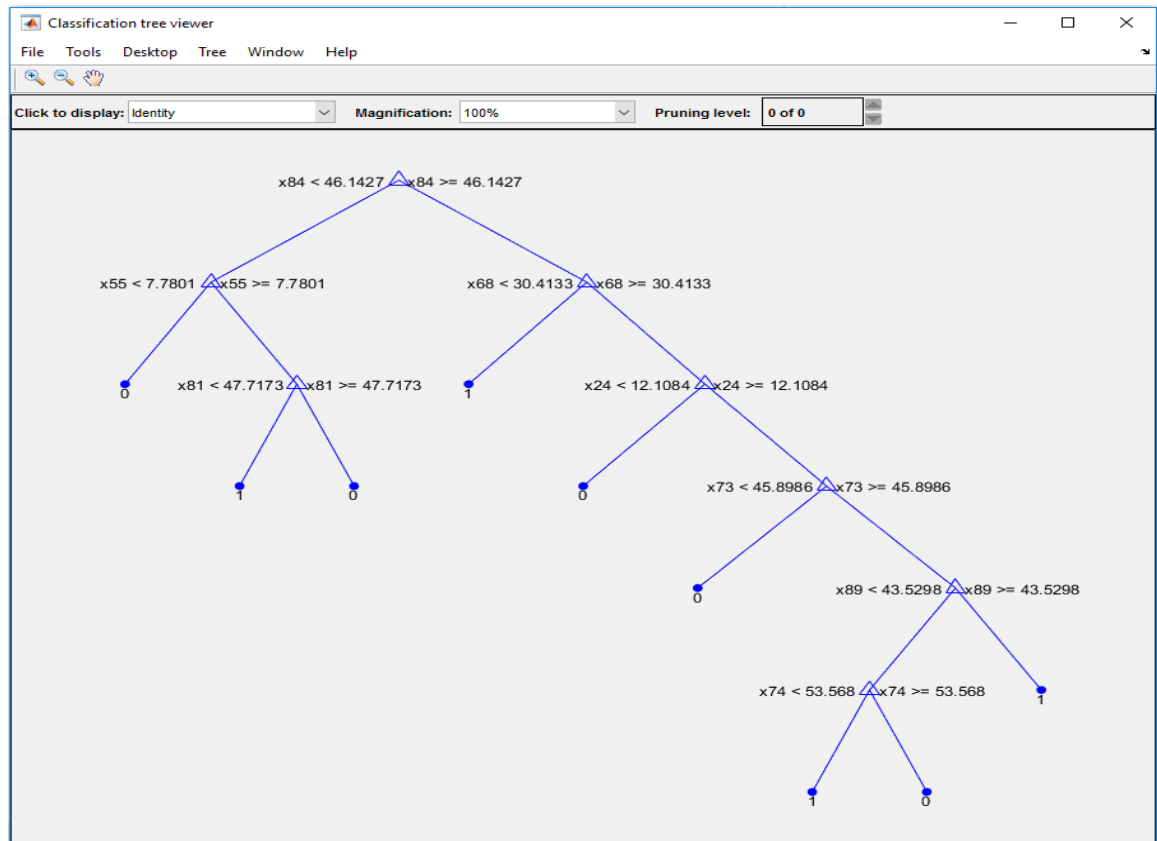
```

Decision tree for classification
1  if x56<12.8733 then node 2 elseif x56>=12.8733 then node 3 else 1
2  if x83<45.8842 then node 4 elseif x83>=45.8842 then node 5 else 0
3  if x35<12.2182 then node 6 elseif x35>=12.2182 then node 7 else 1
4  class = 0
5  if x56<12.0744 then node 8 elseif x56>=12.0744 then node 9 else 1
6  class = 0
7  class = 1
8  if x2<2.63201 then node 10 elseif x2>=2.63201 then node 11 else 1
9  class = 0
10 class = 0
11 if x2<18.2828 then node 12 elseif x2>=18.2828 then node 13 else 1
12 if x25<16.9403 then node 14 elseif x25>=16.9403 then node 15 else 1
13 class = 0
14 if x38<4.93282 then node 16 elseif x38>=4.93282 then node 17 else 1
15 class = 0
16 class = 1
17 if x54<5.65231 then node 18 elseif x54>=5.65231 then node 19 else 1
18 class = 0
19 class = 1

```

**Figure 35 : Description du premier arbre de décision des 10 arbres de la forêt des données de Flexion-extension**





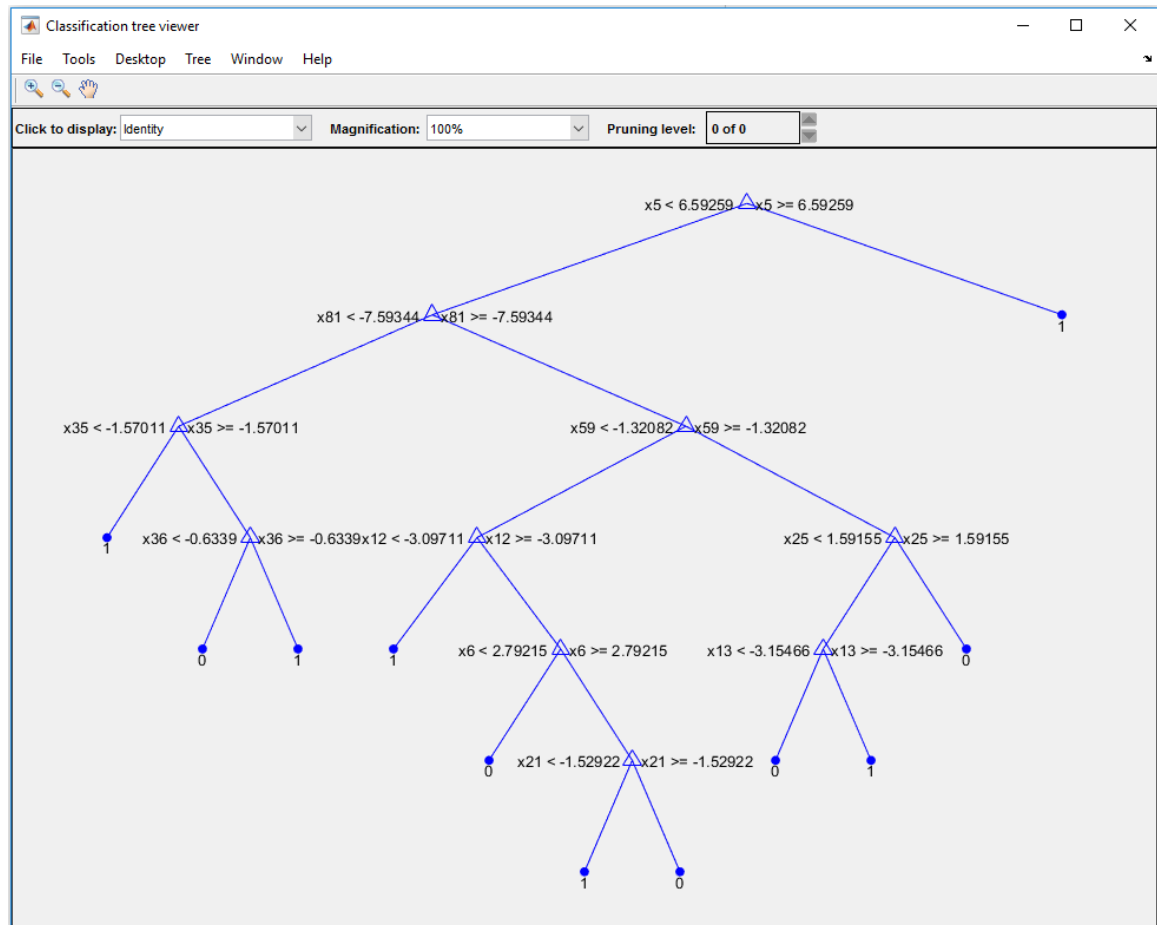
**Figure 36 : Dernier arbre de décision des 10 arbres de la forêt des données de Flexion-extension**

```

Decision tree for classification
1  if x84<46.1427 then node 2 elseif x84>=46.1427 then node 3 else 1
2  if x55<7.7801 then node 4 elseif x55>=7.7801 then node 5 else 0
3  if x68<30.4133 then node 6 elseif x68>=30.4133 then node 7 else 1
4  class = 0
5  if x81<47.7173 then node 8 elseif x81>=47.7173 then node 9 else 0
6  class = 1
7  if x24<12.1084 then node 10 elseif x24>=12.1084 then node 11 else 1
8  class = 1
9  class = 0
10 class = 0
11 if x73<45.8986 then node 12 elseif x73>=45.8986 then node 13 else 1
12 class = 0
13 if x89<43.5298 then node 14 elseif x89>=43.5298 then node 15 else 1
14 if x74<53.568 then node 16 elseif x74>=53.568 then node 17 else 0
15 class = 1
16 class = 1
17 class = 0

```

**Figure 37 : Description du dernier arbre de décision des 10 arbres de la forêt des données de Flexion-extension**



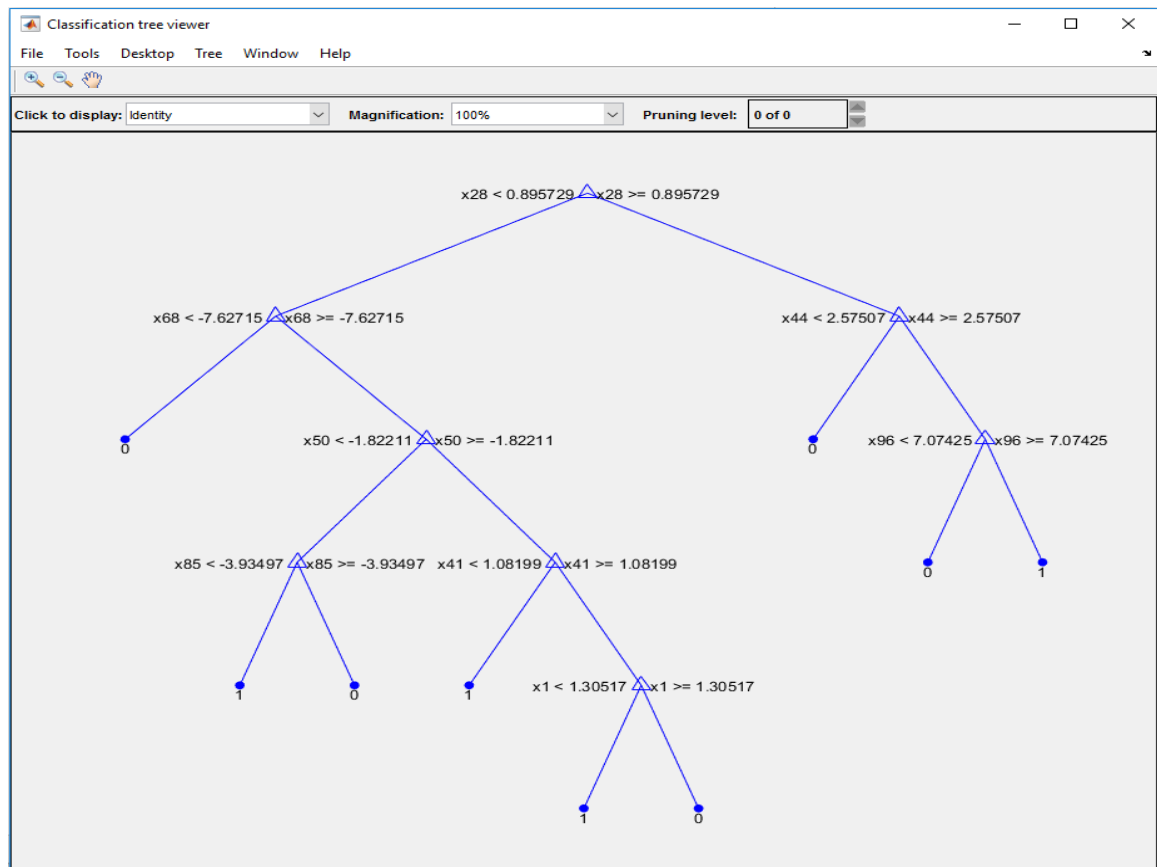
**Figure 38 : Premier arbre de décision des 10 arbres de la forêt des données de Rotation**

```

Decision tree for classification
1  if x5<6.59259 then node 2 elseif x5>=6.59259 then node 3 else 1
2  if x81<-7.59344 then node 4 elseif x81>=-7.59344 then node 5 else 0
3  class = 1
4  if x35<-1.57011 then node 6 elseif x35>=-1.57011 then node 7 else 1
5  if x59<-1.32082 then node 8 elseif x59>=-1.32082 then node 9 else 0
6  class = 1
7  if x36<-0.6339 then node 10 elseif x36>=-0.6339 then node 11 else 0
8  if x12<-3.09711 then node 12 elseif x12>=-3.09711 then node 13 else 0
9  if x25<1.59155 then node 14 elseif x25>=1.59155 then node 15 else 1
10 class = 0
11 class = 1
12 class = 1
13 if x6<2.79215 then node 16 elseif x6>=2.79215 then node 17 else 0
14 if x13<-3.15466 then node 18 elseif x13>=-3.15466 then node 19 else 1
15 class = 0
16 class = 0
17 if x21<-1.52922 then node 20 elseif x21>=-1.52922 then node 21 else 0
18 class = 0
19 class = 1
20 class = 1
21 class = 0

```

**Figure 39 : Description du premier arbre de décision des 10 arbres de la forêt des données de Rotation**



**Figure 40 : Dernier arbre de décision des 10 arbres de la forêt des données de Rotation interne/externe**

```

Decision tree for classification
1  if x28<0.895729 then node 2 elseif x28>=0.895729 then node 3 else 1
2  if x68<-7.62715 then node 4 elseif x68>=-7.62715 then node 5 else 1
3  if x44<2.57507 then node 6 elseif x44>=2.57507 then node 7 else 0
4  class = 0
5  if x50<-1.82211 then node 8 elseif x50>=-1.82211 then node 9 else 1
6  class = 0
7  if x96<7.07425 then node 10 elseif x96>=7.07425 then node 11 else 1
8  if x85<-3.93497 then node 12 elseif x85>=-3.93497 then node 13 else 0
9  if x41<1.08199 then node 14 elseif x41>=1.08199 then node 15 else 1
10 class = 0
11 class = 1
12 class = 1
13 class = 0
14 class = 1
15 if x1<1.30517 then node 16 elseif x1>=1.30517 then node 17 else 0
16 class = 1
17 class = 0

```

**Figure 41 : Description du dernier arbre de décision des 10 arbres de la forêt des données de Rotation interne/externe**

La matrice de confusion est obtenue grâce à l'ensemble de donnée de test. Elle est décrite dans le tableau suivant :

Réels (n = 26)/Prédits	0	1	Total prédit
Non Chirurgicaux	Vrai Positif VP 7	Faux Négatif FN 1	VP + FN 8

Chirurgicaux	Faux Positif FP	Vrai Négatif VN	FP + VN
	10	8	18
Total	VP + FP	FN+VN	n
	17	9	26

**Tableau 5 : Matrice de confusion des données d'abduction-adduction**

La sensibilité : c'est la capacité d'obtenir un résultat positif alors que le sujet est chirurgical, elle est calculée par :

$$S_e = \frac{VP}{VP + FN} = 87.50\%$$

La spécificité : c'est la capacité d'obtenir un résultat négatif alors que le sujet est non chirurgical, elle est calculée par :

$$S_p = \frac{VN}{VN + FP} = 44.40\%$$

Le taux de classification ou la précision: c'est le pourcentage des exemples correctement classés, il est calculé par :

$$T_c = \frac{VP + VN}{VP + VN + FP + FN} = 57.69\%$$

Le taux d'erreur : c'est la quantité de faux positif par rapport à l'ensemble des individus ayant servi à la validation du modèle.

$$T_e = \frac{FP + FN}{VP + VN + FP + FN} = 42.30\%$$

Le taux de faux positif : c'est la quantité d'individus qui sont prédits vrai alors qu'en réalité ils sont faux.

$$T_p = \frac{FP}{FP + VN} = 55.55\%$$



Le taux de faux négatif : c'est la quantité d'individus qui sont prédits faux alors qu'en réalité ils sont vrais.

$$T_n = \frac{FN}{FN + VP} = 12.50\%$$

Pour les données de flexion-extension, suite à la matrice de confusion, on peut calculer les différents taux suivants :

Réels/Prédits	0	1	Total prédit
0	Vrai Positif VP 6	Faux Négatif FN 2	VP + FN 8
1	Faux Positif FP 9	Vrai Négatif VN 9	FP + VN 18
Total	VP + FP 15	FN+VN 11	n 26

**Tableau 6 : Matrice de confusion des données de flexion-extension**

La sensibilité : c'est la capacité d'obtenir un résultat positif alors que le sujet est chirurgical, elle est calculée par :

$$S_e = \frac{VP}{VP + FN} = 75.00\%$$

La spécificité : c'est la capacité d'obtenir un résultat négatif alors que le sujet est non chirurgical, elle est calculée par :

$$S_p = \frac{VN}{VN + FP} = 50.00\%$$

Le taux de classification ou la précision: c'est le pourcentage des exemples correctement classés, il est calculé par :

$$T_c = \frac{VP + VN}{VP + VN + FP + FN} = 57.69\%$$

Le taux d'erreur : c'est la quantité de faux positif par rapport à l'ensemble des individus ayant servi à la validation du modèle.

$$T_c = \frac{FP + FN}{VP + VN + FP + FN} = 42.30\%$$

Le taux de faux positif : c'est la quantité d'individus qui sont prédits vrai alors qu'en réalité ils sont faux.

$$T_p = \frac{FP}{FP + VN} = 50.00\%$$

Le taux de faux négatif : c'est la quantité d'individus qui sont prédits faux alors qu'en réalité ils sont vrais.

$$T_p = \frac{FN}{FN + VP} = 25.00\%$$

Enfin, pour les données de rotation interne-externe, suite à la matrice de confusion, on peut calculer les différents taux suivants :

Réels/Prédits	0	1	Total prédit
Non Chirurgicaux	Vrai Positif VP 5	Faux Négatif FN 3	VP + FN 8
Chirurgicaux	Faux Positif FP 12	Vrai Négatif VN 6	FP + VN 18
Total	VP + FP 17	FN+VN 9	n 26

**Tableau 7 : Matrice de confusion des données de rotation interne-externe**

La sensibilité : c'est la capacité d'obtenir un résultat positif alors que le sujet est chirurgical, elle est calculée par :

$$S_e = \frac{VP}{VP + FN} = 62.50\%$$

La spécificité : c'est la capacité d'obtenir un résultat négatif alors que le sujet est non chirurgical, elle est calculée par :

$$S_p = \frac{VN}{VN + FP} = 33.33\%$$

Le taux de classification ou la précision: c'est le pourcentage des exemples correctement classés, il est calculé par :

$$T_c = \frac{VP + VN}{VP + VN + FP + FN} = 42.30\%$$

Le taux d'erreur : c'est la quantité de faux positif par rapport à l'ensemble des individus ayant servi à la validation du modèle.

$$T_e = \frac{FP + FN}{VP + VN + FP + FN} = 57.69\%$$

Le taux de faux positif : c'est la quantité d'individus qui sont prédits vrai alors qu'en réalité ils sont faux.

$$T_p = \frac{FP}{FP + VN} = 66.67\%$$

Le taux de faux négatif : c'est la quantité d'individus qui sont prédits faux alors qu'en réalité ils sont vrais.

$$T_n = \frac{FN}{FN + VP} = 37.50\%$$

Nous pouvons récapituler le tout dans le tableau suivant :

Base de données	Sensibilité	Spécificité	Taux de classification	Taux d'erreur	Taux de FP	Taux de FN
Abduction - adduction	87.50%	44.44%	57.69%	42.30%	55.55%	12.50%
Flexion - extension	75.00%	50.00%	57.69%	42.30%	50.00%	25.00%
Rotation interne - externe	62.50%	33.33%	42.30%	57.69%	66.67%	37.50%

**Tableau 8 : Récapitulatif des taux mesurés par les matrices de confusion**

Dans le plan frontal (l'abduction et l'adduction), malgré que notre base de données soit limitée, la sensibilité présente un taux très élevé de 87.50%. Ce qui explique la capacité très élevée d'un sujet d'obtenir un résultat positif alors que le sujet est chirurgical tandis que la spécificité et le taux d'erreur tendent tous deux vers la moyenne autrement dit, la probabilité d'obtenir un résultat négatif alors que le sujet est non chirurgical est presque égale à la quantité de faux positif par rapport à l'ensemble des individus ayant servi à la validation du modèle. Les sujets correctement classés présentent un taux de

57.69% contre 12.50% pour les sujets qui sont prédits faux alors qu'en réalité ils sont vrais.

De la même manière, dans le plan transversal (les mouvements de rotation axial), la capacité élevée d'un sujet d'obtenir un résultat positif alors que le sujet est chirurgical présente un taux de 62.50%, la probabilité d'obtenir un résultat négatif alors que le sujet est non chirurgical est 33.33% tandis que la quantité de faux positif par rapport à l'ensemble des individus ayant servi à la validation du modèle est de 57.69% et les sujets correctement classés présentent un taux de 42.30%.

Enfin, dans le plan sagittal (les mouvements de flexion et d'extension), la sensibilité présente un taux très élevé de 75.00%. Ce qui explique la capacité très élevée d'un sujet d'obtenir un résultat positif alors que le sujet est chirurgical tandis que la spécificité et le taux de faux positif présentent un taux moyen de 50.00% enfin la précision ou le taux des sujets correctement classés est de 57.69%.



## Conclusion générale

Ce travail nous a donné l'occasion de découvrir divers domaines du Data Mining, en particulier la classification des formes avec les forêts aléatoires. Ces dernières sont en général plus efficaces que les simples arbres de décision mais possèdent l'inconvénient d'être plus difficilement interprétables. Les arbres de décisions ne fournissent pas uniquement une solution au problème de classification (prédire la survenue d'une maladie en fonction des données de l'individu) mais permet également de retracer les questions à poser pour effectuer ce choix.

Bien que ces arbres possèdent certaines propriétés intéressantes, leur construction est fortement dépendante de l'échantillon initial ce qui peut provoquer des problèmes si certaines données sont fausses ou si la base de données est restreinte. Même si les données sont exactes, il est possible que l'ajout de quelques données à l'échantillon change complètement l'arbre ainsi construit. Nous nous sommes intéressés à étudier la classification des formes avec un modèle ensembliste appelé Random Forest sur une tâche de classification reliée au domaine médical. Afin d'avoir un système performant pour la reconnaissance d'un sujet chirurgical du genou, nous avons implémenté l'algorithme des forêts aléatoires avec pour objectif de minimiser l'erreur et d'effectuer une meilleure classification. Les résultats obtenus en utilisant cet algorithme sont très prometteurs. Néanmoins, Certaines propriétés particulières des forêts aléatoires

soulevées au cours du présent travail méritent un complément d'étude. Nous avons brièvement abordé l'utilisation des forêts aléatoires pour des tâches de classification non supervisée. Pour générer les arbres de décision, les forêts aléatoires utilisent une version du bagging modifiée substantiellement. Comparativement au bagging, les forêts aléatoires sélectionnent un nombre d'attributs aléatoires dans la construction des nœuds de chaque arbre au lieu d'utiliser la totalité des attributs comme le fait le bagging. Par après, un vote de majorité est effectué pour déterminer la classe majoritaire en pondérant uniformément la classification de chaque arbre de la forêt. Les résultats obtenus ne sont pas très compétitifs. La question à laquelle on a essayé donc de répondre était de savoir s'il existe une distribution optimale qui pourrait permettre d'avoir une performance encore meilleure que celle des classifications et des forêts. À partir donc des données paracliniques du patient, nous pouvons générer en sortie un système d'aide à la décision chirurgicale.

## ANNEXES

### 5. Annexe A

```
%-----%

% Chargeons dans Matlab le contenu de la base de données en utilisant
% la commande xlsread.

%-----%

clc;

filename = 'Abduction_Apprentissage_Test.xlsx';

sheet = 2;

xlRange = 'A2:CX64';

DonneesApprentissage = xlsread(filename,sheet,xlRange);

[n m] = size(DonneesApprentissage);

% Déterminons la taille des données à utiliser

Donnees = DonneesApprentissage(:,2:m);

[ligne colonne] = size(Donnees);

% La commande ci-dessous nous permet d'obtenir les mêmes résultats chaque fois que
nous exécutons le code.
```

```

% rng default

% Les 100 premieres colonnes représentent les variables des individus.

% La dernière colonne représente la classe Label (Ce que nous voulons prédire)

trainData = (Donnees);

features = trainData(:,(1:colonne-1));

classLabels = trainData(:,colonne);


% Combien d'arbres aimerions-nous avoir dans notre forêt?

nTrees = 20;

rng(1); % Pour la reproductibilité

B = TreeBagger(nTrees,features,classLabels,'oobPred','On','Method','classification')


view(B.Trees{1})    % Affichage description du premier arbre

view(B.Trees{1},'mode','graph')    % Affichage du premier arbre

view(B.Trees{nTrees}) %Affichage description du dernier arbre

view(B.Trees{nTrees},'mode','graph') %Affichage du dernier arbre

% B.Trees{nTrees}.NumNodes

% A=B.Trees{1}.NumNodes

```

```

%-----
%

% Pour tester une donnée, à présent notre BD sans les étiquettes 0 et 1 afin de calculer
les 3 critères d'évaluation

% Que devrait être la classe Label (0 ou 1)?

%-----
%

% Chargeons la base de données Abduction à tester sans le Label (0 et 1) de Classe
pour la prédiction

% La BD à tester doit avoir les descripteurs suivants:

% - V1 à V100 : Vitesses de la marche des sujets selon les différentes étapes du cycle
de la marche.

%-----
%

TrainDataTest = xlsread(filename,3);

[lig colon] = size(TrainDataTest)

% TrainDataTest = xlsread('FichierBDaTester');

DonneesTest = TrainDataTest(:,(2:colon));

```

```
% % Prédisons la décision de la Forêt formée.
```

```
prediction0ou1B = B.predict(DonneesTest);
```

```
% % La prédiction est un caractère. Nous voulons qu'elle soit un nombre.
```

```
predictedClass = str2double(prediction0ou1B)
```

```
%-----
```

```
%
```

## 6. Annexe B

```
%-----%
```

```
% Chargeons dans Matlab le contenu de la base de données en utilisant
```

```
% la commande xlsread.
```

```
%-----%
```

```
clc;
```

```
filename = 'Flexion_Apprentissage_Test.xlsx';
```

```
sheet = 2;
```

```
xlRange = 'A2:CX64';
```

```
DonneesApprentissage = xlsread(filename,sheet,xlRange);
```

```
[n m] = size(DonneesApprentissage);
```

```
% Déterminons la taille des données à utiliser

Donnees = DonneesApprentissage(:,2:m);

[ligne colonne] = size(Donnees);

% La commande ci-dessous nous permet d'obtenir les mêmes résultats chaque fois que
% nous exécutons le code.

% rng default

% Les 100 premières colonnes représentent les variables des individus.

% La dernière colonne représente la classe Label (Ce que nous voulons prédire)

trainData = (Donnees);

features = trainData(:,(1:colonne-1));

classLabels = trainData(:,colonne);

% Combien d'arbres aimerions-nous avoir dans notre forêt?

nTrees = 10;

rng(1); % Pour la reproductibilité

B = TreeBagger(nTrees,features,classLabels,'oobPred','On','Method','classification')

view(B.Trees{1}) % Affichage description du premier arbre
```

```

view(B.Trees{1},'mode','graph') % Affichage du premier arbre

view(B.Trees{nTrees}) %Affichage description du dernier arbre

view(B.Trees{nTrees},'mode','graph') %Affichage du dernier arbre

% B.Trees{nTrees}.NumNodes

% A=B.Trees{1}.NumNodes

%-----

%

% Pour tester une donnée, à présent notre BD sans les étiquettes 0 et 1 afin de calculer
les 3 critères d'évaluation

% Que devrait être la classe Label (0 ou 1)?

%-----

%

% Chargeons la base de données Abduction à tester sans le Label (0 et 1) de Classe
pour la prédiction

% La BD à tester doit avoir les descripteurs suivants:

% - V1 à V100 : Vitesses de la marche des sujets selon les différentes étapes du cycle
de la marche.

%-----

%
```



```

TrainDataTest = xlsread(filename,3);

[lig colon] = size(TrainDataTest)

% TrainDataTest = xlsread('FichierBDaTester');

DonneesTest = TrainDataTest(:,(2:colon));

% % Prédisons la décision de la Forêt formée.

prediction0ou1B = B.predict(DonneesTest);

% % La prédiction est un caractère. Nous voulons qu'elle soit un nombre.

predictedClass = str2double(prediction0ou1B)

%-----

%

```

## 7. Annexe C

```

%-----%

% Chargeons dans Matlab le contenu de la base de données en utilisant

% la commande xlsread.

%-----%

clc;

filename = 'Rotation_Apprentissage_Test.xlsx';

sheet = 2;

```

```
xlRange = 'A2:CX64';

DonneesApprentissage = xlsread(filename,sheet,xlRange);

[n m] = size(DonneesApprentissage);

% Déterminons la taille des données à utiliser

Donnees = DonneesApprentissage(:,2:m);

[ligne colonne] = size(Donnees);

% La commande ci-dessous nous permet d'obtenir les mêmes résultats chaque fois que
% nous exécutons le code.

% rng default

% Les 100 premières colonnes représentent les variables des individus.

% La dernière colonne représente la classe Label (Ce que nous voulons prédire)

trainData = (Donnees);

features = trainData(:,(1:colonne-1));

classLabels = trainData(:,colonne);

% Combien d'arbres aimerions-nous avoir dans notre forêt?

nTrees = 10;

rng(1); % Pour la reproductibilité
```

```

B = TreeBagger(nTrees,features,classLabels,'oobPred','On','Method','classification')

view(B.Trees{1})    % Affichage description du premier arbre

view(B.Trees{1},'mode','graph')    % Affichage du premier arbre

view(B.Trees{nTrees}) %Affichage description du dernier arbre

view(B.Trees{nTrees},'mode','graph') %Affichage du dernier arbre

% B.Trees{nTrees}.NumNodes

% A=B.Trees{1}.NumNodes

%-----

%

% Pour tester une donnée, à présent notre BD sans les étiquettes 0 et 1 afin de calculer
les 3 critères d'évaluation

% Que devrait être la classe Label (0 ou 1)?

%-----

%

% Chargeons la base de données Abduction à tester sans le Label (0 et 1) de Classe
pour la prédiction

% La BD à tester doit avoir les descripteurs suivants:

% - V1 à V100 : Vitesses de la marche des sujets selon les différentes étapes du cycle
de la marche.

```

```
%-----  
  
%  
  
TrainDataTest = xlsread(filename,3);  
  
[lig colon] = size(TrainDataTest)  
  
% TrainDataTest = xlsread('FichierBDaTester');  
  
DonneesTest = TrainDataTest(:,(2:colon));  
  
% % Prédisons la décision de la Forêt formée.  
  
prediction0ou1B = B.predict(DonneesTest);  
  
% % La prédiction est un caractère. Nous voulons qu'elle soit un nombre.  
  
predictedClass = str2double(prediction0ou1B)  
  
%-----  
  
%
```

## Bibliographie

Youssef OUAKRIM, 2011 Classification de sujets asymptomatiques et gonarthrosiques en fonction des données cinématiques : comparaison de l'approche globale et de l'approche locale

Anderson, J. J., et D. T. Felson. 1988. « Factors associated with osteoarthritis of the knee in the first national Health and Nutrition Examination Survey (HANES I). Evidence for an association with overweight, race, and physical demands of work ». *Am J Epidemiol*, vol. 128, no 1 (Jul), P. 179-189.

Arden, N., et M. C. Nevitt. 2006. « Osteoarthritis: epidemiology ». *Best Pract Res Clin Rheumatol*, vol. 20, no 1 (Feb), P. 3-25.

Allard P., Blanchi J.P. 1996 « Analyse du mouvement humain par la biomécanique », Ed. Ville Mont-Royal, 281 p.

Barbier F. 1994 « Modélisation biomécanique du corps humain et analyse de la marche normale et pathologique – Application à la rééducation », Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambrésis, Valenciennes

Breiman, L. et al., *Classification and regression trees*, Chapman & Hall, New York, 1984.

Breiman, L., « Heuristics of instability and stabilization in model selection », *Ann. Stat.*, 24(6), p. 2350–2383, 1996a.

Breiman, L., « Bagging predictors », *Machine Learning*, 24, p. 123–140, 1996b.

Breiman, L., Bias, variance, and arcing classifiers, rapport technique, 1996c.

Breiman, L., « Random forests », *Machine Learning*, 45, p. 5–32, October 2001.

L. Breiman. Bagging predictors. *Machine Learning*, 24(2), 1996.

Gérard Biau. Analysis of a random forests model. 2010.

L. Breiman. Random forests. *Machine Learning*, 45, 2001.

Brenière Y., Do M.C. 1991 « Control of gait initiation », *Journal of Motor Behaviour*, Vol. 23, n° 4, p. 235-240

Buckland-Wright, C. 2006. « Which radiographic techniques should we use for research and clinical practice? ». *Best Pract Res Clin Rheumatol*, vol. 20, no 1 (Feb), P. 39-55.

Boulic R., Thalmann N.M., Thalmann D. (1990) « A global human walking model with real-time kinematic personification », *The Visual Computer*, Vol. 6, n° 6, p. 344-358

Cicuttini, F. M., T. Spector et J. Baker. 1997. « Risk factors for osteoarthritis in the tibiofemoral and patellofemoral joints of the knee ». *J Rheumatol*, vol. 24, no 6 (Jun), P. 1164-1167.

Astephen, Janie L., Kevin J. Deluzio, Graham E. Caldwell, Michael J. Dunbar et Cheryl L. Hubley-Kozey. 2007. « Gait and neuromuscular pattern changes are associated with

differences in knee osteoarthritis severity levels ». *Journal of Biomechanics*, vol. 41, no 4, P. 868-876.

Whittle M.W. 1996 « Clinical gait analysis: A review », *Human Movement Science*, Vol. 15, n° 3, p. 369-387

Whittle, Michael. 2003. *Gait analysis : an introduction*, 3rd. Edinburgh ; New York: Butterworth-Heinemann, x, P. 220

Whittle M. 2002. *Gait analysis : an introduction*. Butterworth-Heinemann Medical.

Williams M., Lissner H.R., Le Veau B.F. (1986) « Biomécanique du mouvement humain, une introduction », Mount Royal, Québec: Décarie Éditeur, 2e Ed.

Plas F., Viel E., Blanc Y. (1989) « La marche humaine. Kinésiologie dynamique. Biomécanique et pathomécanique », 2e Ed. Paris: Masson, 122 p.

Viel E. 2000 « La marche humaine, la course et le saut : biomécanique, exploration, normes et dysfonctionnements », *Le point en rééducation et en APS*, Paris: Masson, 267 p.

Kapandji I.A. 1996 « Physiologie articulaire, Tome 2 : membre inférieur », Maloine, 5e Ed, 384 p.

Calais-Germain B. 1986 « Anatomie pour le mouvement tome 1 : Introduction à l'analyse des techniques corporelles », Limoux, France, Ed. Calais-Germain B.

Chao E.Y.S. 1980 « Justification of triaxial goniometer for measurement of joint rotation », *Journal of Biomechanics*, Vol. 13 p. 989-1006

El Hafi F., Gorce P. (1999) « Behavioural approach for a bipedal robot stepping motion gait », *Robotica*, Cambridge University Press, Vol. 17, n° 5, p. 491-501

El Hafi F. 2000 « Modélisation dynamique du mouvement : application à la marche et à l'enjambement d'obstacle », Thèse de doctorat, Université Paris XII Val de Marne, Paris

Faivre A. 2003 « Conception et validation d'un nouvel outil d'analyse de la marche », Thèse de doctorat, Université de Franche-Comté, Franche-Comté

Felson, D. T. 1995. « Weight and osteoarthritis ». *J Rheumatol Suppl*, vol. 43, (Feb), P. 7-9.

Freund, Y., Schapire, R.E.. Experiments with a new boosting algorithm. In *Proceedings of the 13th International Conference on Machine Learning*, p.148-156, 1996.

Gaitosik R.L., Bohannon R.L. (1987) « Clinical measurement of range of motion/ Review of goniometry emphasizing reliability and validity », *Physical Therapy*, Vol. 67, p. 1867-1872

Gillet C. 2004 « Analyse biomécanique de la marche et proposition de classes de marcheurs - Application au portage de sacs à dos », Thèse de doctorat, Université de Valenciennes et du HainautCambrésis, Valenciennes



Gorce P. 2000 « Approche biomécanique et robotique du mouvement chez l'homme. »

Mémoire d'habilitation à diriger des recherches, Université Paris Sud Orsay, Paris

Gorce P., Guihard M. (1998) « On dynamic control of pneumatic biped », Journal of Robotic Systems, Vol. 15, n° 7, p. 421-433

Gorce P., Boughiden M.L. (2003) « Biomechanical modeling of « seat to stand process » on electrical wheelchair », AMSE the international journal of « Modeling, Measurement, Control », Part C , supplement 2C, Vol. 63, p. 58-71

Goujon H. 2006 « Analyse de la marche de l'amputé fémoral », Thèse de Doctorat, ENSAM, Paris

Hart, D. J., et T. D. Spector. 1993. « The relationship of obesity, fat distribution and osteoarthritis in women in the general population: the Chingford Study ». J Rheumatol, vol. 20, no 2 (Feb), P. 331-335.

Hunter, DJ. 2009. « Focusing osteoarthritis management on modifiable risk factors and future therapeutic prospects ». Therapeutic Advances in Musculoskeletal Disease, no 1, P. 35-47.

Inman V., Ralston H., Todd F. et Lieberman J. (1981). Human walking. Williams & Wilkins Baltimore.

Jones, L., M. J. Beynon, C. A. Holt et S. Roy. 2006. « An application of the Dempster-Shafer theory of evidence to the classification of knee function and detection of

improvement due to total knee replacement surgery ». *J Biomech*, vol. 39, no 13, p. 2512-2520.

Jones, L., C. A. Holt et M. J. Beynon. 2008. « Reduction, classification and ranking of motion analysis data: an application to osteoarthritic and normal knee function data ». *Comput Methods Biomech Biomed Engin*, vol. 11, no 1 (Feb), p. 31-40.

Lawrence, R. C., D. T. Felson, C. G. Helmick, L. M. Arnold, H. Choi, R. A. Deyo, S. Gabriel, R. Hirsch, M. C. Hochberg, G. G. Hunder, J. M. Jordan, J. N. Katz, H. M. Kremers et F. Wolfe. 2008. « Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II ». *Arthritis Rheum*, vol. 58, no 1 (Jan), P. 26-35

Milne A.D., Chess D.G., Johnson J.A., King G.J.W. (1996) « Accuracy of an electromagnetic tracking device: A study of the optimal operating range and metal interference », *Journal of Biomechanics*, Vol. 29, p.791-793

Multon F. 1998 « Contrôle du mouvement des humanoïdes de synthèse », Thèse de doctorat, Université de Rennes I, Rennes

Padgaonkar A.J., Krieger K.W., King A.I. (1975) « Measurement of angular acceleration of a rigid body using linear accelerometers », *Journal of Applied Mechanics*, Vol. 42, p. 552-556

Pélissier J. 1997 « La marche chez l'hémiplégique: bilan et modalité thérapeutique », *Les troubles de la marche et leur exploration*, Ed. Frison-roche

Pélissier J., Boisson D. (1994) « Les paramètres de la marche humaine techniques actuelles d'exploration », La marche humaine et sa pathologie, Paris ed. Masson, p. 41-55

Perry J. 1992. Gait analysis : normal and pathological function. SLACK incorporated.

Slobounov S.M., Simon R.F., Bush J.A., Kraemer W.J., Sebastianelli W., Slobounova E. (1999) « An alternate method of range of motion assessment », Journal of Strength and Conditioning Research, Vol. 13 , n° 4, p. 389-393

Srikanth, V. K., J. L. Fryer, G. Zhai, T. M. Winzenberg, D. Hosmer et G. Jones. 2005. « A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis ». Osteoarthritis Cartilage, vol. 13, no 9 (Sep), P. 769-781.

Vapnik, V.N. The nature of statistical learning theory. Springer Verlag, 2000.

Wang X. 1999 « A behavior-based inverse kinematics algorithm to predict arm prehension postures for computer-aided ergonomic evaluation », Journal of Biomechanics, Vol. 32, n° 5, p. 453-460

Watelain E. 1999 « Proposition d'une méthode d'analyse du mouvement : Application à la marche du sujet hémiplégique », Thèse de doctorat Université de Lille 2, Lille

Neila Mezghani, TÉLUQ, 2015 Note de cours Forage de données, Module 2 Analyse en composantes principale

Neila Mezghani, TÉLUQ, 2015 Note de cours Forage de données, Module 3  
Regroupement

Neila Mezghani, TÉLUQ, 2015 Note de cours Forage de données, Module 4 Arbre de  
décision

Neila Mezghani, TÉLUQ, 2015 Note de cours Forage de données, Module 5  
Corrélation et régression

N. Mezghani, N. Gaudreault, A. Fuentes, A. Mitiche, N. Hagemeister, R. Aissaoui, et  
J.A. de Guise Identification des patrons de la cinématique du genou par une analyse en  
composantes principales Identification of knee kinematic patterns in normal gait by  
principal component analysis

Youssef OUAKRIM, 2011 Classification de sujets asymptomatiques et  
gonarthrosiques en fonction des données cinématiques : comparaison de l'approche  
globale et de l'approche locale

## Webographie

1. [https://www.cihi.ca/fr/types-de-soins/services-specialises/arthroplasties#\\_1](https://www.cihi.ca/fr/types-de-soins/services-specialises/arthroplasties#_1)
2. <http://www.doctissimo.fr/html/sante/atlas/fiches-corps-humain/squelette-genou.htm>
3. <http://www.bio-top.net/Terminologie/F/femur.htm#Fémur>
4. [http://techniquerollin.com/mecanique\\_jambe.php](http://techniquerollin.com/mecanique_jambe.php)
5. <http://amar-constantine.e-monsite.com/pages/anatomie-physiologie/osteologie.html>
6. <http://www.physiotherapiepour tous.com/anatomie/anatomie-du-genou/>
7. <http://www.physiotherapiepour tous.com/douleur-au-genou/fissure-du-menisque-ou-lesion/>
8. <http://clinique-jouvenet-paris.ramsaygds.fr/vous-%C3%AAtes-patient-pourquoi-choisir-notre-%C3%A9tablissement-tous-nos-soins/arthrose-du-genou-26>
9. <https://www.medacta.com/ch-fr/all/patients-and-caregivers/knee/le-genou-et-la-gonarthrose>
10. <http://www.douleurgenou.fr/arthrose/arthrose-chronique.php>
11. <http://www.monarthrose.ca/definition-arthrose-du-genou/>
12. <https://www.arthrocoach.com/fr/tout-sur-arthrose/diagnostic-arthrose>